Saturday, October 19, 10:00 AM - 12:30 PM

Summit - 448

D-Wave Quantum Optimization

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Quantum Optimization

Catherine Potts, D-Wave Quantum Inc, Palo Alto, CA, United States

Quantum computing has gone from the lab to the enterprise, and a recent Hyperion Research study reported that respondents overwhelmingly indicated that quantum computing-based optimization capabilities are seen as integral to optimizing their key business processes. While you may think production use of quantum computers are years away, the first commercial quantum applications in production are using D-Wave's quantum technology. This workshop will cover the following topics: An introduction to quantum computers, quantum annealing, and optimization solvers What kinds of business problems are good candidates for quantum/ hybrid solutions Use case examples: how quantum hybrid is being used for optimization applications such as resource and workforce scheduling, cargo loading and other complex business problems How quantum systems are programmed, and the tools for developers. This will include live demos and an introduction to Leap, D-Wave's real-time cloud service. Leap provides access to a portfolio of hybrid solvers, enabling enterprises to address all kinds of business problems that range in size and complexity. How to get started with quantum for business.

Saturday, October 19, 1:00 PM - 3:30 PM

Summit - 447

GAMSPy and Data APIs for Streamlining Optimization

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - GAMSPy and Data APIs for Streamlining Optimization

Atharv Bhosekar, GAMS Development Corp, Fairfax, VA, United States

GAMS (General Algebraic Modeling System) is an algebraic modeling language that provides users a way to write optimization models using intuitive algebraic syntax. However, as optimization becomes an integrated step within larger decision-making pipelines, modelers face two significant challenges: (1) the inconvenience of switching out of a preferred programming language (such as Python) solely for optimization purposes, and (2) the difficulty of efficiently transferring data between GAMS and other tools and platforms within a diverse software ecosystem. In this presentation, we will tackle these challenges using our latest solutions. First, we will present GAMSPy, our new product that brings algebraic modeling capabilities to Python. GAMSPy allows users to enjoy an intuitive algebraic syntax without compromising on the performance. We will also highlight our suite of data APIs to streamline data exchange with GAMS. In particular, we will focus on GAMS Transfer, a data API that enables users of R, MATLAB, and Python to efficiently read, modify, analyze, and write GAMS data. These tools significantly enhance the interoperability of GAMS within multi-platform decision pipelines, facilitating smoother and more efficient optimization workflows.

Summit - 445

Elevate your Optimization Practice with latest Advances in FICO® Xpress Solver and Xpress Insight

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Elevate your Optimization Practice with latest Advances in FICO® Xpress Solver and Xpress Insight

Dinakar Gade, FICO, Bozeman, MT, United States, Michael Perregaard, Leona Gottwald, Tristan Gally, Majid Bazrafshan, Bruno Vieira, Alexander Biele

Join our workshop to discover the cutting-edge innovations in Xpress Solver 9.5 and our new object-oriented APIs! Get a firsthand look at how Xpress Insight empowers business users to harness the full potential of analytical models faster than ever before.

Experience the latest features and performance enhancements in FICO® Xpress Solver, including groundbreaking advances in mixed-integer linear and nonlinear optimization. These include a new, first-order hybrid gradient algorithm for linear optimization problems, new heuristics, cutting and branching techniques. Furthermore, we discuss updates to Xpress Global to solve general mixed-integer nonlinear, and nonconvex problems to proven global optimality.

The new FICO Xpress Solver API leads to a dramatically reduced overall model building time and increased memory efficiency. It supports modern programming concepts such as Collections, Streams, Lambdas, and operator overloading. All in an interface that guarantees a consistent user experience across different programming languages.

The session will cover advancements in the Python API, outlining recent updates to modelling capabilities and discussing best practice for developing readable and yet efficient implementations of large-scale optimization models.

FICO® Xpress Insight is a rapid application development and deployment framework that integrates with Xpress Solver and your own analytics. We will show how you can rapidly convert Python models into complete business applications with Xpress Insight to make your analytical models available to thousands of business users.

Summit - 446

Analyzing Multidimensional Data and building Predictive Models in an Interactive No Code Environment

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Analyzing Multidimensional Data and building Predictive Models in an Interactive No Code Environment

Kevin Potcner, JMP Statistical Discovery, San Francisco, CA, United States

Fully analyzing data requires an analyst to generate a wide variety of visualizations, analyses, and models. The insights gleaned from each step leads the analyst to a next set of analyses to try, and so on. This exploratory approach can become very cumbersome in a coding environment. Without strong proficiency in programming, students often get frustrated finding themselves unable to run many different analyses and visualization quickly. In this presentation, a statistical scientist from JMP will illustrate how easy it can be to explore multi-dimensional data through interactive visualizations, analyses, and prediction models in a “no code” environment. The audience will see how this approach not only greatly expedites data analysis efforts, but provides students with a much richer and more engaging experience to learning analytics. The content presented is most relevant to: Professional (Mid-Career); Associate (Early Career); and Executive (Senior Level).

Summit - 448

Hexaly, a New Kind of Global Optimization Solver

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - HEXALY, A New Kind of Global Optimization Solver

Fred Gardi, Hexaly, Brooklyn, NY, United States

Hexaly is a new kind of global optimization solver. Its modeling interface is nonlinear and set-oriented. It also supports user-coded functions, thus enabling black-box optimization and, more particularly, simulation optimization. In a sense, Hexaly APIs unify modeling concepts from mixed-linear programming, nonlinear programming, and constraint programming. Under the hood, Hexaly combines various exact and heuristic optimization methods: spatial branch-and-bound, simplex methods, interior-point methods, augmented Lagrangian methods, automatic Dantzig-Wolfe reformulation, column and row generation, propagation methods, local search, direct search, population-based methods, and surrogate modeling techniques for black-box optimization. Regarding performance benchmarks, Hexaly distinguishes itself against the leading solvers in the market, like Gurobi, IBM Cplex, and Google OR Tools, by delivering fast and scalable solutions to problems in the space of Supply Chain and Workforce Management like Routing, Scheduling-Packing-Clustering-Matching-Assignment-Location problems. For example, on notoriously hard problems like the Pickup/Delivery Problem with Time-Windows or Flexible Job Shop-Scheduling with Setup Times, Hexaly delivers solutions with a gap to the best solutions known in the literature smaller than 1% in a few minutes of running times on a basic computer. In addition to the Optimizer, we provide an innovative development platform called Hexaly Studio to model and solve rich Vehicle-Routing and Job-Shop Scheduling problems in a no-code fashion. The user can define its problem and data, run the Optimizer, visualize the solutions and key metrics through dashboards, and deploy the resulting app in the cloud – without coding. The content is most relevant to: Associate (Early Career), Professional (Mid-Career), and Executive (Senior Level).

Summit - 444

Linking Logistics Models: Combining ML and OR Through Ensembling and Horizontal Computing

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Linking logistics models: Combining ML and OR through ensembling and horizontal computing

Carolyn Mooney, Nextmv.io Inc, Philadelphia, PA, United States

Conceptually, the connection between machine learning (ML) and operations research (OR) is intuitive: order demand predictions feed into scheduling delivery operations, consumer buying behavior influences inventory management, and so on. But in practice, seamlessly blending the two disciplines can be more challenging than expected due to technical implementation and the ability to manage uncertainty. This interactive session will explore foundational ML and OR logistics models (forecasting, scheduling, routing, etc.) alongside the tooling and techniques to link, solve, and orchestrate them for real-world, operational settings. We will demonstrate ways to manage uncertainty, employ ensembling to aggregate multiple forecasts, run multiple model scenarios, and derive consensus across model output to feed downstream applications. Join us for an interactive session that will explore foundational logistics models, best practices for operations such as model testing and CI/CD, and tooling such as OR-Tools, Pyomo, HiGHS, Gurobi, AMPL, Statsmodels, Prophet, and more. This content is most relevant to Associate (Early Career); Professional (Mid-Career); and Executive (Senior Level).

Saturday, October 19, 4:00 PM - 6:30 PM

Summit - 447

Introducing ChiAha – the Smart Manufacturing Digital Twin Toolkit

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Introducing ChiAha – the Smart Manufacturing Digital Twin Toolkit

Amber Siprelle, Chiaha.ai, Tallassee, TN, United States

ChiAha can predict production line performance and OEE within 1% accuracy. High-fidelity modeling constructs with statistically modeled data-driven behavior. Provide answers to many of the questions related to the design, operation and improvement of lines for optimum OEE. This content is most relevant to Associate (Early Career); Professional (Mid-Career); and Executive (Senior Level).

Summit - 444

What's New in Gurobi 12.0: Helping You to Build, Solve, and Deploy Optimization Models

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - What's New in Gurobi 12.0: Helping You to Build, Solve, and Deploy Optimization Models

Xavier Nodet, Gurobi Optimization, LLC, Antibes, France, Gregory Glockner, Ed Klotz, Rodrigo Fuentes, Zed Dean, Maliheh Aramon

Our pre-conference workshop will cover a wide range of updates and practical tips, including a preview of new features that will be available with Gurobi 12.0, an overview of our new massive open online course (MOOC), and an interactive panel on GenAI and Optimization. Gurobi 12.0 Overview: This talk will provide a preview of notable features in the upcoming Gurobi 12.0 release. Besides the usual performance comparisons against previous releases, we will consider advances in the global MINLP solver that was first introduced in the current version, 11.0. These include direct support for compound multivariate nonlinear expressions that eliminate the need for users to create auxiliary variables and constraints. Major advances in the efficiency of the Web License Server will also be discussed, as well as other features that improve the usability of the optimizer. A Groundbreaking MOOC: We'll also be presenting a short overview of "Introduction to Optimization Through the Lens of Data Science," a groundbreaking massive-open online course (MOOC) developed by Gurobi in partnership with Dr. Joel Sokol, professor at Georgia Tech. This course provides a unique opportunity for anyone to enhance their skill sets and for educators to bring cutting-edge, practical knowledge into their classrooms. Panel–Generative AI and Optimization: The workshop will conclude with a panel discussion featuring optimization experts from across industries, academia, and Gurobi as they discuss generative AI and optimization. All attendees will receive a special Gurobi t-shirt. The content presented is most relevant to: Associate (Early Career); Professional (Mid-Career); and Executive (Senior Level).

Summit - 448

Littlefield 2.0: A New Version of the Online Game for Operations Management Courses

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - Littlefield 2.0: A New Version of the Online Game for Operations Management Courses

Samuel Wood, Responsive Learning Techologies, Los Altos, CA, United States

After 25 years there is a new version of Littlefield! Littlefield is a competitive online simulation of either a factory or a medical laboratory that has been by more than half a million students in 500+ universities in 60+ countries to excite and engage students in operations management topics like process analysis and inventory control. This presentation will introduce a newly updated version 2 of the game that was released this past summer. Participants are encouraged but not required to bring a laptop. This content is most relevant to: Executive (Senior Level); Professional (Mid-Career); and Associate (Early Career).

Summit - 446

SAS Analytics

Invited Session

Exhibitor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Dr Ste 200, Catonsville, MD, 21228, United States

1 - SAS Analytics

Yan Xu, SAS Institute, Cary, NC, United States, Rob Pratt, Laci Ladanyi, Bochuan Lyu

SAS offers extensive analytic capabilities, including machine learning, deep learning, natural language processing, statistical analysis, optimization, and simulation. SAS analytic functionality is also available through the open, cloud-enabled design of SAS® Viya®. You can program in SAS or in other languages – Python, Lua, Java, and R. SAS Analytics is equipped with AI-enabled automations and modern low-code or no-code user interfaces that democratize data science usage in your organization and offer unparalleled speed to value. We will first review the SAS analytics portfolio, then highlight recently added optimization features (including automated Benders decomposition and support for Pyomo and PuLP), and finally explore several case studies in optimization. The content presented is most relevant to Associate (Early Career); Professional (Mid-Career); and Executive (Senior Level).

Sunday, October 20, 8:00 AM - 9:15 AM

SA01

Summit - 320

Healthcare Data Analytics

Invited Session

Service Science

Chair: Mohammad Fili, Oklahoma State University, Stillwater, OK, United States

1 - Improving Menu Selection in Assisted Living Facilities by Considering Residents' Behavior

Sadan Kulturel-Konak, Penn State Berks, Reading, PA, United States

The proportion of the US population residing in assisted living facilities is steadily rising. Often, elderly individuals in these facilities depend solely on the meals provided to fulfill their nutritional requirements. Studies indicate that many seniors in assisted living have deficits in several crucial macro- and micronutrients, alongside overconsumption of sodium, fat, cholesterol, and saturated fat relative to their age demographic. This study is centered on devising nutritious menus for assisted living facilities while juggling multiple objectives and intricate nutritional limitations. The proposed approach considers a range of factors in menu development, encompassing USDA healthy eating guidelines, input from chefs based on their expertise and preferences, menu diversity, cost considerations, and the varying dietary needs and preferences of residents. One significant advancement over prior research lies in integrating individuals' menu selection behaviors, thereby enhancing the menus' practicality upon implementation and fostering a sense of autonomy among residents.

2 - Pain Quantification: A Novel Decision-Support System Benefitting from Multi-Attribute Decision-Making Techniques and Large Language Models

Mohammad Alipour Vaezi, Virginia Tech, BLACKSBURG, VA, United States, Kwokleung Tsui, Huaiyang Zhong

Historically, various approaches have been employed for pain assessment. Nevertheless, none of the existing methodologies can capture the comprehensive extent of the patient's pain due to their intrinsic limitations. In this research, for the first time in history, a novel approach is introduced for pain quantification considering various aspects of pain gathered from different sources to determine the real pain level of the patients suffering from acute or chronic pains. In this research, first, each pain assessment technique is reviewed, and its limitations are identified. Then, a novel method is conducted by integrating two Multi-Attribute Decision-Making (MADM) Techniques (Bayesian Best-Worst Method and Weighted aggregated sum product assessment) based on experts' opinions. The experts in this study include both humans (physicians) and artificial intelligence (Chat GPT 4 and Google Gemini). To validate the proposed method, it is applied to two different groups of patients suffering from acute (group I) and chronic (group II) pain as a real-life case study. The results indicate the reliability of the proposed approach for both groups. This research provides an interactive tool that can be helpful for every healthcare system in which patients are suffering from acute or chronic pain.

3 - GANCQR: Estimating Prediction Intervals for Individual Treatment Effects with GANs

Jiaxing Wang, North Carolina State University, Raleigh, NC, United States, Hong Wan, Xi Chen

Evaluating individual treatment effects (ITE) poses challenges due to the lack of access to counterfactuals, especially when dealing with biased data. Currently, there is a growing interest in harnessing the generative capabilities of Generative Neural Networks (GANs) and Variational Autoencoders (VAEs) for ITE estimation. However, few methods effectively tackle uncertainty quantification of the estimated ITE. In this work, we introduce GANCQR, a GAN-based conformal prediction approach aimed at providing prediction intervals of ITE with acceptable coverage. Numerical evaluations on synthetic and semi-synthetic datasets underscore GANCQR's superiority in handling datasets with selection bias compared to state-of-the-art methods.

SA02

Summit - 321

QSR Best Referred Paper

Award Session

Quality, Statistics and Reliability

Chair: Ashif Iquebal, Arizona State University, Tempe, AZ, United States

Co-Chair: Xiaochen Xian, University of Florida, Gainesville, FL, United States

Co-Chair: Mostafa Reisi, University of Florida, Gainesville, FL, United States

1 - 4DYNAMO: Analyzing and Optimizing Process Parameters in 4D Printing for Dynamic 3D Shape Morphing Accuracy

Michael Biehler, Georgia Institute of Technology, Atlanta, GA, United States, Daniel Lin, Reinaldo Mock, Jianjun Shi

Additive manufacturing (AM), commonly referred to as 3D printing, has undergone significant advancements, particularly in the realm of stimuli-responsive 3D printable and programmable materials. This progress has led to the emergence of 4D printing, a fabrication technique that integrates AM capabilities with intelligent materials, introducing dynamic functionality as the fourth dimension. Among the stimuli-responsive materials, shape memory polymers have gained prominence, notably for their crucial applications in stress-absorbing components. However, the exact 3D shape morphing of 4D printed products is affected by both the 3D printing conditions as well as the stimuli activation. Hence it has been hard to precisely control the 3D shape morphing accuracy. To model and optimize the dynamic 3D evolution of the 4D printed parts, we conducted both simulation studies and real-world experiments and introduced a novel machine-learning approach extending the concept of normalizing flows. This method not only enables the process optimization of the dynamic 3D profile evolution by optimizing the process conditions during 3D printing and stimuli activation but also provides interpretability for the intermediate shape morphing process. This research contributes to a deeper understanding of the nuanced interplay between process parameters and the dynamic 3D transformation process in 4D printing.

2 - Multi-physics Guided Generative Diffusion Models with Manufacturing Applications

Naichen Shi, University of Michigan, Ann Arbor, MI, United States, Hao Yan, Shenghan Guo, Raed Al Kontar

coming soon

3 - Real-time Transfer Active Learning for Functional Regression and Prediction based on Multi-output Gaussian Process

Zengchenghao Xia, University of Iowa, Iowa City, IA, United States, Zhiyong Hu, Qingbo He, Chao Wang

Active learning provides guidance for the design and modeling of systems with highly expensive sampling costs. However, existing active learning approaches suffer from cold-start concerns, where the performance is impaired due to the initial few experiments designed by active learning. In this paper, we propose using transfer learning to solve the cold-start problem of functional regression by leveraging knowledge from related and data-rich signals to achieve robust and superior performance, especially when only a few experiments are available in the signal of interest. More specifically, we construct a multi-output Gaussian process (MGP) to model the between-signal functional relationship. This MGP features unique innovations that distinguish the proposed transfer active learning from existing works: i) a specially designed covariance structure for characterizing within-and between-signal inter-relationships and facilitating interpretable transfer learning, and ii) an iterative Bayesian framework to update the parameters and prediction of the MGP in real-time, which significantly reduces the computational load and facilitates the iterative active learning. The inter-relationship captured by this novel MGP is then fed into active learning using the integrated mean-squared error (IMSE) as the objective. We provide theoretical justifications for this active learning mechanism, which demonstrates the IMSE is monotonically decreasing as we gather more data. The real-time updating and the monotonically decreasing objective together provide both practical efficiency and theoretical guarantees for solving the cold-start concern in active learning. The proposed method is compared with benchmark methods through various numerical and real case studies, and the results demonstrate the superiority of the method.

4 - Spatial In-Profile Monitoring via Latent Tensor-variate Gaussian Process with Mixed Effects

Peiyao Liu, Tsinghua University, Beijing, China, People's Republic of, Chen Zhang

Nowadays advanced sensing technology enables real-time data collection in two or higher-dimensional coordinate systems, known as spatial profiles. These data have attracted significant efforts toward anomaly detection and quality control in manufacturing. However, most of the existing monitoring methods face detection delays as they require complete profiles before implementation. To accommodate the sequential nature of the data, Gaussian Process (GP) modeling offers a promising approach. Yet, the high-dimensional covariance matrix associated with densely sampled spatial profiles pose challenges in estimation accuracy. Additionally, they often treat spatial locations as the GP inputs, neglecting numerous latent factors involved in the generation of spatial profiles. This article proposes an in-profile monitoring (INPOM) control chart for spatial profiles based on a latent tensor-variate GP with mixed effects (LTGP-ME) model. The random effects component is constructed by a LTGP, which preserves the multiway structure of spatial profiles in the tensor domain while capturing underlying nonlinear variations through latent factors. Fixed effects component can be further incorporated to capture the relationship between spatial profiles and input variables. We develop an expectation maximization algorithm for parameter estimation, exploring model identifiability and convergence properties. Based on the prediction errors of LTGP-ME, a Hotelling T^2 statistic is further constructed for INPOM. The effectiveness and applicability of the proposed approach are demonstrated through extensive simulation studies and a real case study in additive manufacturing.

SA03

Summit - 322

Knowledge Fusion for Modeling and Decision-making in Complex Systems

Invited Session

Quality, Statistics and Reliability

Chair: Shancong Mou, Georgia Institute of Tehnology/University of Minnesota Twin Cities, Minneapolis, MN, United States

1 - Local Transfer Gaussian Process for Surrogate Modeling of Expensive Computer Simulators

XINMING WANG, Peking University, Beijing, China, People's Republic of, John Miller, Jianguo Wu, Simon Mak

A critical bottleneck for scientific progress is the costly nature of computer simulations for complex systems. Surrogate models provide an appealing solution: such models are trained on simulator evaluations, then used to emulate and quantify uncertainty on the expensive simulator at unexplored inputs. In many applications, one often has available data on related systems. For example, in designing a new jet turbine, there may be existing studies on turbines with similar configurations or geometries. A key question is how to transfer information from such ``source" systems for effective surrogate training on the ``target" system. We thus propose a new LOcal transfer Learning Gaussian Process (LOL-GP) model, which leverages a carefully-designed GP to transfer such information for surrogate modeling. The key novelty of the LOL-GP is the use of a data-learned regularization model, which learns regions where transfer should be performed and where it should be avoided. This ``local transfer" property is desirable in scientific systems: for certain parameters, the systems may behave similarly and thus transfer is beneficial; for others, the systems may be considerably different and thus transfer is detrimental. In accounting for local transfer, the LOL-GP can remedy an important limitation of ``negative transfer" in existing transfer learning models, where the transfer of information worsens predictive performance. We derive a Gibbs sampling algorithm for efficient posterior predictive sampling on the LOL-GP. We then show, via a suite of numerical experiments and an application to jet turbine design, the improved surrogate performance of the LOL-GP over the existing state-of-the-art.

2 - Uni-3Dad: Gan-Inversion Aided Universal 3D Anomaly Detection on Model-Free Products

Jiayu Liu, Rensselaer Polytechnic Institute, Troy, NY, United States, Shancong Mou, Yinan Wang

Anomaly detection is a long-standing challenge in manufacturing systems, which aims to identify the location of surface defects and critically impact product quality. 3D point clouds have started to attract practitioners’ attention due to their robustness to environmental factors and

capability of geometric representation. The existing 3D anomaly detection lies in two branches. One line of work directly compares the scanned 3D point cloud with the design file, which assumes the design file is always available. However, there are many model-free products, such as fresh produce (i.e., cookies, bagels, potatoes), dentures, etc. The other line of work is to compare the patches of scanned 3D point clouds with a dictionary of normal patches. However, when detecting incomplete shapes of the target product (i.e., missing pieces of cookie and denture), there is no representation of the missing regions in the collected 3D point cloud due to the nature of 3D scanning, which fails the existing methods of identifying such type of anomaly. To resolve these two challenges, we proposed a unified, unsupervised 3D anomaly detection framework for model-free products. Specifically, GAN-inversion is firstly adapted in the 3D anomaly detection to accurately identify the missing regions. In addition, the One-class Support Vector Machine is tailored to fuse the detection results from feature-based and GAN-inversion-based branches. The results demonstrated that (1) our proposed method receives comparable performance with the state-of-the-art (SOTA) methods on detecting anomalies except for incomplete shapes and (2) it outperforms the SOTA methods on identifying incomplete shapes.

3 - Classifying Phase Behavior of Soft Materials using Spectral Data

Xuerui Song, Virginia Tech, Blacksburg, VA, United States

Understanding soft material composition-phase-property relations is essential for predicting material processability and part quality, such as for additive manufacturing and biomanufacturing applications using hydrogels. This work is driven by the challenge of predicting the composition-phase-mechanical property relation data for hydrogels using spectral data generated by autonomous experimentation (AE) workflows. While AE workflows can efficiently generate spectral data, the data quality and characteristics are material- and application-dependent. We hypothesize that deep learning can help address the challenges of classifying spectral data that exhibit variance in signal-to-noise ratio, sparsity, and class imbalance, which are common to practical experimental applications. Here, we examine the performance of a one-dimensional Convolutional Neural Network (1D-CNN) for understanding and predicting the phase behavior of sustainable α-cyclodextrin (α-CD) hydrogels using impedance spectra data from piezoelectric cantilever sensors. Spectral data dimensionality was first reduced using principal component analysis (PCA) followed by clustering algorithms, identifying the spectral data labels and class distribution (i.e., solid, transition, or liquid states). The unsupervised learning studies indicated that α-CD hydrogels exhibited a third class associated with a 'transition' region in the gel mechanical properties. Supervised learning with a 1D-CNN model outperformed traditional non-deep methods with respect to prediction accuracy and precision. This work highlights the applicability and effectiveness of 1D-CNN models to overcome challenges of classifying spectral data generated in practical sensing and AE-driven materials science and chemistry applications and advances our understanding of the composition-phase-property behavior of a sustainable hydrogel

4 - Spatio-temporal modeling and its applications to extreme environmental processes

Guanzhou Wei, Georgia Institute of Technology, Atlanta, GA, United States, Xiao Liu

In this presentation, we showcase two wildfire applications using the advanced spatio-temporal models. In the first application, motivated by the significant impacts of wildfire aerosols on solar energy production, we propose a physics-informed statistical modeling approach for wildfire aerosol propagation using heterogeneous satellite remote sensing data streams. The model has successfully integrated multi-source remote sensing data streams with the underlying physical advection-diffusion process and is able to handle the heterogeneity among multi-source data streams (e.g., missing data, systematic biases, etc.). In the second application, to quantify the wildfire risks for the power delivery infrastructures, we develop a new spatio-temporal point process model, known as the Convolutional Non-homogeneous Poisson Process (cNHPP) on linear networks. This model incorporates both the topology of power delivery networks and the cumulative effects of real-time environmental factors, surpassing the performance of conventional NHPP models.

5 - Multistage Net: Learning Continuous Multistage Manufacturing Processes of Liquid Products without Intermediate Output and Lead Time Labels

Chiehyeon Lim, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea, Republic of

Manufacturers have implemented the continuous multistage manufacturing processes (MMPs), in which raw materials are processed through multiple stages without interruptions, based on efficiency and flexibility compared to discrete MMPs. While there are numerous methods of using machine learning to estimate conditions and outputs of manufacturing processes, applying this approach to continuous MMPs of liquid products is challenging for the following reasons: First, obtaining intermediate output labels is very difficult because the process is originally designed to operate without interruptions. Second, determining lead times between stages is also difficult for liquid products, which are mixed between stages during production. To address these challenges, we propose "Multistage Net," a novel machine learning model designed for continuous MMPs of liquid products. Multistage Net is composed of several "interstage blocks" organized in a hierarchical structure within a "multistage module", with these modules being stacked in the model. The interstage block captures sequential dependency between previous and current stages without requiring intermediate output and concurrently explores lead-time relationship via attention and gated mechanisms. The multistage module effectively learns the sequential nature of MMPs across all stages without requiring intermediate outputs through interconnected interstage blocks. Validation studies on liquid sugar and lube base oil manufacturing with real-world datasets demonstrate the superior performance of Multistage Net compared to existing models, showing its effectiveness in both production yield forecasting and quality prediction tasks. Further analyses not only confirm the exclusive utilities of components within Multistage Net but show it delivers impressive performance without need for lead time labels.

SA04

Summit - 323

Advanced Data Analytics for IoT-Enabled Smart and Connected Systems

Invited Session

Quality, Statistics and Reliability

Chair: Xubo Yue, Northeastern University, Boston, MA, United States

Co-Chair: Seokhyun Chung, University of Virginia, Charlottesville, VA, 22903, United States

1 - Federated Automatic Latent Variable Selection in Multi-output Gaussian Processes

Jingyi Gao, University of Virginia, Charlottesville, VA, United States, Seokhyun Chung

This paper explores a federated learning approach that automatically selects the only necessary number of latent processes in multi-output Gaussian processes (MGPs). The MGP has seen great success as a transfer learning tool when data is generated from multiple sources/units/entities. A common approach in MGPs to transfer knowledge across units involves gathering all data from each unit to a central server and extracting common independent latent processes to express each unit as a linear combination of the shared latent patterns. However, this approach poses key challenges in (i) determining the adequate number of latent processes and (ii) centralized learning which leads to potential privacy risks and significant computational burdens on the central server. To address these issues, we propose a hierarchical model that places spike-and-slab priors on the coefficients of each latent process. These priors help automatically select only needed latent processes by shrinking the coefficients of unnecessary ones to zero. To facilitate federated inference, we propose a variational inference-based approach, formulating model inference as an optimization problem compatible with federated learning algorithms. This approach allows units to jointly select and infer the common latent processes without sharing their data. We also discuss an efficient learning approach for a new unit within our proposed federated framework. Simulation and the case studies on Li-ion battery capacity degradation and air temperature data demonstrate the advantageous features of our proposed approach.

2 - A General Second-Order Augmented Primal-Dual Framework for Distributed Optimization

Runxiong Wu, University of Wisconsin, Madison, Madison, WI, United States, Andi Wang

We develop a general second-order augmented primal-dual method for distributed optimization, specifically designed for problems where the objective function is composed of a strongly convex twice differentiable term and a possibly non-differentiable convex regularizer. Our approach leverages the proximal operator of the regularizer to transform the augmented Lagrangian of the dual problem into a novel minimax optimization problem, making it highly suitable for distributed optimization algorithms. By incorporating a generalization of the Hessian, we define second-order updates that enhance the convergence rates and solution quality. We demonstrate the efficiency and robustness of our method, even when dealing with non-smooth regularizers such as the ℓ_1 penalty. To improve practical applicability, we employ a line search strategy to automatically tune the step sizes, which incurs no additional communication overhead. Extensive experiments on real distributed datasets validate our framework, showing significant performance improvements over state-of-the-art methods. Our results highlight the framework's ability to handle complex optimization problems efficiently, making it a valuable tool for researchers and practitioners in distributed optimization.

3 - Domain-Informed Reinforcement Learning for Multi-Component Preventive Maintenance Planning under Economic Dependence

Jaesung Lee, Texas A&M University, College Station, TX, United States, Tatthapong Srikitrungruang, Salman Jahani

Many engineering systems consist of multiple components working simultaneously. These components undergo gradual and irreversible degradation, leading to failure, at varying rates. Preventive maintenance is essential to mitigate these failures and reduce costs.
In multi-component systems, maintenance actions are economically interdependent, influenced by factors such as operational shutdowns and shared labor, resulting in high setup costs.
Our goal is to identify an optimal preventive maintenance policy that minimizes total costs in multi-component systems with economic dependence, where each component degrades stochastically at randomly varying speeds.
Preventive maintenance in multi-component systems is challenging, as the action space grows exponentially with the number of components, further complicated by economic dependencies.
Existing decision-making methods, including deep reinforcement learning, fall short in addressing such challenges.
Naturally, most studies have considered a single or a few components in a system for maintenance planning.
To bridge this gap, we introduce a domain-informed reinforcement learning approach, integrating domain knowledge via a sparsity-enhanced stochastic hierarchical policy and reward shaping with Bayesian degradation predictions.
Our carefully designed policy function successfully characterizes the sporadic nature of maintenance events.
Extensive numerical studies and a real-world case study on frying machines demonstrate the significant benefits and substantial cost savings of our method.

SA05

Summit - 324

INFORMS JFIG Best Paper Competition II

Award Session

Junior Faculty Interest Group

Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Incentivizing Resource Pooling

Chen Chen, New York University Shanghai, Shanghai, China, People's Republic of

Resource pooling improves system efficiency drastically in large stochastic systems, but its effective implementation in decentralized systems remains relatively underexplored. This paper studies how to incentivize resource pooling when agents are self-interested, and their states are private information. Our primary motivation is applications in the design of decentralized computing markets, among others. We study a standard multi-server queueing model in which each server is associated with an M/M/1 queue and aims to minimize its time-average job holding and processing costs. We design a simple token-based mechanism where servers can earn tokens by offering help and spend tokens to request help from other servers, all in their self-interest. The mechanism induces a complex game among servers. We employ the fluid mean-field equilibrium (FMFE) concept to analyze the system, combining mean-field approximation with fluid relaxation. This framework enables us to derive a closed-form characterization of servers' FMFE strategies. We show that these FMFE strategies approximate well the servers' rational behavior. We leverage this framework to optimize the design of the mechanism and present our main results: As the number of servers increases, the proposed mechanism incentivizes \emph{complete} resource pooling---that is, the system dynamics and performance under our mechanism match those under centralized control. Finally, we show that our mechanism achieves the first-best performance even when helping others incurs higher job processing costs and remains nearly optimal in settings with heterogeneous servers.

2 - Sample Complexity of Inventory Control with Setup Costs Beyond Plug-in Demand Estimation

Xiaoyu Fan, Stern School of Business, New York University, New York, NY, United States

We show in this work that a class of structured MDPs admits more efficient learning (i.e., lower sample complexity bounds) compared to the best possible/known algorithms in generic RL. We focus on the MDPs describing the inventory control system with fixed ordering costs, a fundamental problem in supply chains. We develop an algorithm applied to the inventory MDPs, which leads to strictly lower sample complexity bounds compared to the optimal or best-known bounds recently obtained for the general MDPs. We improve on those ``best-possible'' bounds by carefully leveraging the structural properties of the inventory dynamics in various settings.

3 - Designing Payments Models for the Poor

Bhavani Shanker Uppari, Singapore Management University, Singapore, Singapore, Sasa Zorc

Several basic services, such as energy, clean water and cooking gas, are currently out of reach for millions of people living in poverty. There has been an emergence of private firms that offer these services by, for example, selling solar home systems or clean cooking packages with remote lockout capabilities. These firms deploy a pay-as-you-go model in which consumers are given the flexibility to manage the amount and frequency of their payments based on their own erratic cash flows. However, because a firm under this model cannot observe how much money the consumers have, they can pay less to the firm and turn their income to other needs. We employ an optimal contracting approach to investigate the incentives that can mitigate such misuse of payment flexibility.

SA06

Summit - 325

Location Analysis for Electric Vehicle Charging

Invited Session

Location Analysis

Chair: Ibrahim Capar, University of North Carolina Wilmington, Bowling Green, OH, United States

1 - Improving Disadvantaged Communities by Locating EV Chargers

Ibrahim Capar, University of North Carolina Wilmington, Wilmington, NC, United States, Michael Lash, Ozgur Araz

In this research, we investigate some of the characteristics of disadvantaged communities defined by the “Justice 40” initiative. Justice 40 is a US Federal Program to improve communities around the US by allocating forty percent of a certain Federal Investment. By analyzing historical data, we explore the impact of setting up EV charging locations on disadvantaged communities. Finally, we formulate a mathematical model to maximize EV coverage while selecting disadvantaged communities.

2 - Location-allocation problem with semi-discrete optimal transport

Nanshan Chen, The Ohio State University, Columbus, OH, United States, Cathy Xia, Guzin Bayraksan

Solving facility location-allocation problems in modern transportation settings can be computationally challenging when the demand is dense. We formulate the allocation problem as a semi-discrete optimal transport (OT) problem under a continuous demand density. We then develop a decomposition algorithm combined with Stochastic Gradient Descent (SGD) to solve the two-stage location-allocation problem. We demonstrate the efficiency of our approach in the context of micro-transit services. The applications can be easily extended to EV charging stations and E-Scooter rental stations.

3 - Integrating Intermediate and Destination Charging to Enhance Electrification for Lake Michigan Circuit Tourism Trips

Amirali Soltanpour, Michigan State University, East Lansing, MI, United States, Alireza Rostami, Mehrnaz Ghamami, Ali Zockaie

This study focuses on developing electric vehicle (EV) charging infrastructure within the Lake Michigan area, aiming to provide convenient and efficient charging options, particularly for tourists. The primary objective is to devise an integrated, cost-effective, and convenient charging network to enhance ecotourism within this region. To achieve this, a framework is proposed to integrate and optimize both Level-2 and Direct Current Fast Charging (DCFC) networks. The Level-2 network planning incorporated a 4-step modeling approach including estimations of EV and energy demand, and charger allocation. Conversely, the DCFC network planning considers charging buffer ranges and an algorithm for assigning EVs to charging stations, stochastic queuing delays, and realistic charging behavior for tourism purposes. The Level-2 problem is solved using both mixed-integer linear programming and heuristic approaches, while the DCFC solution approach is proposed based on simulated annealing, dynamic penalty methods, and two-stage decision-making techniques. The study explores intricate trade-offs between capital investment and user costs, and reveals that budget constraints significantly affect infrastructure allocation and user experience. The study suggests dedicating a minimum of 30% of the budget to the DCFC network to maintain an acceptable level of service. An optimal allocation of 60% of the budget to Level-2 charging and 40% to DCFC can effectively meet energy demands and proves to be the most cost-efficient configuration. Overall, the study offers actionable guidance for creating a resilient and user-focused charging infrastructure in the Lake Michigan Circuit area, contributing to the region’s sustainable development and tourism industry growth.

4 - CAST: A Correlation-based clustering-Assisted Sparse auToregressive model based on Lasso

Muting Ma, University of Alabama, Tuscaloosa, AL, United States, Mesut Yavuz, Matthew Hudnall, Qin Wang

We address an electric vehicle (EV) adoption prediction model for charging station planning. Based on the Least Absolute Shrinkage and Selection Operator (Lasso), the prediction model can accurately estimate future EV adoption at a core-based statistical area (CBSA) granularity level. One of the predictors is the public charging station density (#/square miles). Given the historical density and the predicted

EV adoption, future charging station planning can be advised. We test the spatial discrepancy between the expected station density and the density of newly constructed ones in 2023.

5 - Optimizing Urban Infrastructure Planning for Electric Vehicles Considering Stochastic User Charging Behavior

Alireza Rostami, Michigan State University, East Lansing, MI, United States, Ali Azockaie, Mehrnaz Ghamami, Amirali Soltanpour

The rapid adoption of electric vehicles (EVs) presents a promising avenue for reducing fossil fuel dependency and mitigating the environmental impacts of transportation systems. However, many major cities still lack adequate charging infrastructure to support the mass adoption of EVs for urban trips. Existing research on charging infrastructure planning often oversimplifies user charging behavior by assuming charging events occur only when a user's state of charge drops below a minimum threshold, overlooking the complexities and stochastic nature of user charging patterns. This study proposes an innovative approach to model the spatiotemporal decision-making of EV users during public charging and assess its impact on optimal charging infrastructure configuration. Since urban travel is usually a part of trip chains, EV users can have a wide range of options on where and when to charge their vehicles. This study considers dynamic factors such as charging price, trip urgency, proximity to the station, and state of charge to influence users' decision-making process. The study adopts a solution approach based on simulated annealing to minimize charging station placement cost and EV users' time and cost spent during public charging. To model users' daily activities, trip schedules, and traffic flows, this research adopts the agent-based transportation model of POLARIS. The proposed framework is applied to the Chicago regional area network, considering various EV ownership and charging behavior scenarios. The results of this study provide valuable guidance for policymakers in designing efficient charging infrastructure that captures users' charging behavior to support the widespread adoption of electric vehicles.

SA07

Summit - 327

AI in Healthcare and Medical Applications

Flash Session

Contributed

Chair: Tong Zhao, Shunde Building, Tsinghua University, Beijing 100084, 100084

1 - Robotic Sorting Systems: Robot Management and Layout Design Optimization

Tong Zhao, Tsinghua University, Beijing, China, People's Republic of, Xi Lin, Fang He, Hanwen Dai

In the contemporary logistics industry, automation plays a pivotal role in enhancing production efficiency and expanding industrial scale. Autonomous mobile robots, in particular, have become integral to the modernization efforts in warehouses. One noteworthy application in robotic warehousing is the robotic sorting system (RSS), distinguished by its characteristics such as cost-effectiveness, simplicity, scalability, and adaptable throughput control. While previous research has focused on analyzing the efficiency of RSS, it often assumed an ideal robot management system ignoring potential queuing delays by assuming constant travel times. This study relaxes this assumption and explores the quantitative relationship between RSS configuration parameters and system throughput. We introduce a novel robot traffic management method, named the rhythmic control for sorting scenario (RC-S), for RSS operations, equipped with an estimation formula establishing the relationship between system performance and configurations. Simulations validate that RC-S reduces average service time by 10.3% compared to the classical cooperative A* algorithm, while also improving throughput and runtime. Based on the performance analysis of RC-S, we further develop a layout optimization model for RSS, considering RSS configuration, desired throughput, and costs, to minimize expenses and determine the best layout. Numerical studies show that at lower throughput levels, facility costs dominate, while at higher throughput levels, labor costs prevail. Additionally, due to traffic efficiency limitations, RSS is well-suited for small-scale operations like end-of-supply-chain distribution centers.

2 - An Examination of Artificial Intelligence Technologies for Healthcare

Shuyu Lai, Westwood High School, Austin, TX, United States

By leveraging vast amounts of data and efficient learning methods, artificial intelligence (AI) equips machines with human-like intelligence and problem-solving capabilities. Within the healthcare industry, AI presents unique opportunities to assist medical professionals in various aspects, including interacting with patients, collecting information, interpreting medical images, diagnosing medical conditions, and proposing treatment plans. This study delves into these roles, offering a comprehensive review of the latest AI applications and research in healthcare. It demonstrates AI's potential to enhance both efficiency and quality in healthcare delivery, while also highlighting avenues for further research and practical implementation.

3 - RecoCheC: Enhancing Employee Health with Personalized Check-Up Recommendations via Machine Learning

Siwapol Techaratsami, Carnegie Mellon University, Pittsburgh, PA, United States, Saranya Thongsawaeng, Jidapa Hanvoravongchai, Napatsorn Thewaran, Piyawat Kantagowit, Krit Pongpirul

RecoCheC represents a breakthrough in personalized healthcare, leveraging advanced machine learning algorithms to optimize health check-up recommendations based on historical health data. As healthcare costs continue to rise and the demand for personalized medical care increases, RecoCheC offers a strategic, data-driven solution to enhance the effectiveness and efficiency of health assessments.

Utilizing a robust dataset collected from mobile health services over a six-year period (2016-2022), which includes 7,518 records of 3,198 individuals (average 3.4 follow-up years, age 39.3±9.6 years old, male 29.3%) across seven different business entities, RecoCheC integrates demographic and biometric data with historical health outcomes. This innovative platform prioritizes health check-up items through the Individual Historical Lab Score (IHLS), demonstrably outperforming traditional methods based solely on prevalence data or randomized approaches.

The efficacy of RecoCheC was rigorously tested using a Receiver Operating Characteristic (ROC) curve, achieving an impressive Area Under the Curve (AUC) of 0.82. This performance significantly surpasses that of the prevalence model (AUC = 0.77) and the baseline model (AUC = 0.50), underscoring RecoCheC's superior capability in tailoring health check-up schedules to individual needs, thereby potentially reducing unnecessary procedures and focusing resources more effectively.

RecoCheC not only promises significant benefits for individual users but also offers substantial implications for healthcare policy by suggesting a scalable model for reducing overall healthcare expenditures while improving outcomes. The platform's application extends beyond individual health management to include optimizing corporate health strategies, marking a significant advance in the application of machine learning in healthcare.

4 - Generalized Machine Learning Models for Forecasting Extended Length of Stay in Hospitals: Insights from Nationwide Data

Marzieh Amiri Shahbazi, RIT, Rochester, NY, United States, Nasibeh Azadeh Fard

In hospitals, the length of time patients stays (known as "length of stay" or LOS) is an important health outcome metric. LOS can help to analyze the patient’s level of illness, healthcare resource usage, and hospital efficiency. Typically, hospitals categorize patients into cohorts based on their illness or treatment to estimate how long they will stay. However, this method can make it difficult to accurately predict LOS for everyone. Therefore, we aim to develop a new approach that can predict LOS effectively for all patients, regardless of their circumstances. Using different machine learning techniques, we analyze a nationwide dataset with various patient information, such as age, demographics, clinical characteristics, comorbidities, treatment procedures, and hospital-specific factors. By incorporating this information, our goal is to develop and validate a generalized LOS prediction model.

5 - IVExplorer: Interactive Exploration of Causality within Causal Graphs

Ye Ji Chun, Seoul National University, Seoul, Korea, Republic of, Hyunwoo Park

The paper presents an IVExplorer, an interactive visualization system specialized for exploring the causal effect between a given dependent and independent variable. Finding causal effects within a regression equation is an extremely frequent as well as crucial task in many business, economic, and social analysis. One of the most popularly used methods to investigate the causal effect between social variables is using an instrumental variable, which aims to capture the causal effect that x has on y, is assumed to be related with y only through the effect of the independent variable x and only aims to capture the effect x has on y. However, according to Linear Structural Causal Models in the field of causal inference, the causal effect that variable x has on y can be derived through the regression coefficient equation depending on the causal graph of the covariates. The IVExplorer allows the user to create various causal graphs which generates a regression model that calculates the causal effect of variable x on y given the causal graph. By incorporating the causal inference theory to investigate causal effects in social or business analysis, we introduce an intuitive and interactive interface to support decision making for social analytics.

SA08

Summit - 328

OR in Security & Defense I (Mission Support)

Invited Session

Military and Security

Chair: Paul Goethals, Department of Defense, Columbia, MD, United States

1 - Deep Ritz Method with Fourier Feature Mapping: a Deep Learning Approach for Solving Variational Models of Microstructure

Ensela Mema, Kean University, Union, NJ, United States, Ting Wang, Jaroslaw Knap

This work presents a novel approach that combines the Deep Ritz Method (DRM) with Fourier feature mapping to solve minimization problems comprised of multi-well, non-convex energy potentials. These problems present computational challenges as they lack a global minimum. Through an investigation of three benchmark problems in both 1D and 2D, we observe that DRM suffers from spectral bias pathology, limiting its ability to learn solutions with high frequencies. To overcome this limitation, we modify the method by introducing Fourier feature mapping. This modification involves applying a Fourier mapping to the input layer before it passes through the hidden and output layers. Our results demonstrate that Fourier feature mapping enables DRM to generate high-frequency, multiscale solutions for the benchmark problems in both 1D and 2D, offering a promising advancement in tackling complex non-convex energy minimization problems.

2 - Neural Network Cryptography

Lubjana Beshaj, Army Cyber Institute at West Point, West Point, NY, United States

A number of research papers have been published using the architecture of adversarial neural networks to prove that communication between two neural nets based on synchronized input can be achieved, and without knowledge of this synchronized information these systems cannot be breached. In this paper we will try to evaluate these adversarial neural net architectures when a third party gains access to a partial or noisy secret key, has knowledge about a loss function or loss values, or activation functions used during the training of encryption layers. We explore the cryptanalysis side of it in which we will focus on vulnerabilities a neural net-based cryptography network can face. This can be used in the future to improve the current neural net-based cryptography architectures. In this paper we show that while the encryption key is necessary to decrypt the messages in a neural network domain, the adversarial neural networks can occasionally decrypt messages or raise a concern which will require further training.

3 - Bayesian Networks for Predicting Cyber Actor Behaviors

Alicia Bridges, ITA International, LLC, Dayton, MD, United States, Dennis Buede, Charles Burdick

Imagine having a crystal ball to tell you what an adversary's next move might likely be, so that you could prepare to defend against those possible actions. Far too many adversaries target unsuspecting organizations, businesses and other entities with vulnerabilities in their IT networks and are successful at gaining access. Possibly, this is followed by exfiltration of data or worse, launching ransomware attacks that bring their operations to a screeching halt. However, the exploits of many of the cyber actors can be studied as a pattern of behaviors (human or otherwise) and modeled to identify the likely, often repeated steps employed. The ordered attack steps can be arranged into a Bayesian network, which is a probabilistic model where the steps are nodes connected by edges that represent their conditional probabilities via a directed acyclic graph. Through Bayesian network analysis, we can explore modeling the step-by-step patterns of attack behaviors of cyber adversaries and use observed trends in real attacks as a proxy for the probability of those steps occurring and sometimes, given explicit conditions. The steps themselves are represented by the tactics, techniques and procedures (TTPs) that adversaries perform. Through this type of exploratory analysis, entities can use predictive analysis, following the principals of Bayesian analysis, to prepare for predicted cyber behaviors and better defend their networks.

4 - Text Document Triage Using Semantic Embedding, Clustering and Topic Labeling

Kyle Bender, USAF, Baltimore, MD, United States

In this research, we delve into the exploration of advancements in the field of document semantic embedding and clustering techniques. These techniques are pivotal in understanding and interpreting large volumes of data, particularly in the context of sentiment analysis. We further investigate the role of topic labeling, a crucial component in the organization and categorization of data, which aids in the extraction of meaningful insights. The application of these advanced techniques allows for the creation of a hierarchical, two-dimensional visual representation, essentially a canvas that visually maps and organizes data. This visual representation serves as a powerful tool for efficient data analytics, enabling quicker and more accurate interpretation of complex data sets. Our exploration is aimed at broadening the usage of these techniques, with a particular focus on their application in military analysis.

5 - Resilience Measurement for Security Organizations

Paul Goethals, Department of Defense, Columbia, MD, United States

The study of organizational resilience brings with it a host of complexities in derivation, application, interpretation, and perhaps most of all, measurement. The concept has grown significantly in popularity over the last 25 years. Achieving greater resilience is tied closely to increased sustainability and durability; for defense-minded organizations, this translates to higher levels of survivability and adaptability, as well as diminished degrees of risk. This work examines the problem of measuring organizational resilience either in quantitative or qualitative terms through an investigation of various models in the literature. An argument is then presented for ‘why’ challenges exist in establishing such measurement frameworks. Alternative methods for measuring organizational resilience are subsequently offered with explanations on use and their interpretation. The observations and findings tied to this study may be used to further inform efforts in decision analysis or metrics formulation where resilience is a primary attribute of interest.

SA09

Summit - 329

Selected Auctions & Market Design papers from the EC 2024 conference III

Invited Session

Auctions and Market Design

Chair: Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Ali Makhdoumi, Duke University, Durham, NC, United States

Co-Chair: Pengyu Qian, Boston University, Boston, United States

1 - Optimal Stopping with Interdependent Values

Divyarthi Mohan, Boston University, Tel Aviv University, Boston, MA, United States

We study online selection problems in both the prophet and secretary settings, when arriving agents have interdependent values. In the interdependent values model, introduced in the seminal work of Milgrom and Weber [1982], each agent has a private signal and the value of an agent is a function of the signals held by all agents. Results in online selection crucially rely on some degree of independence of values, which is conceptually at odds with the interdependent values model. For prophet and secretary models under the standard independent values assumption, prior works provide constant factor approximations to the welfare. On the other hand, when agents have interdependent values, prior works in Economics and Computer Science provide truthful mechanisms that obtain optimal and approximately optimal welfare under certain assumptions on the valuation functions. We bring together these two important lines of work and provide the first constant factor approximations for prophet and secretary problems with interdependent values. We consider both the algorithmic setting, where agents are non-strategic (but have interdependent values), and the mechanism design setting with strategic agents. All our results are constructive and use simple stopping rules.

2 - Pareto-Optimal Algorithms for Learning in Games

Eshwar Ram Arunachaleswaran, University of Pennsylvania, Philadelphia, PA, United States

We study the problem of characterizing optimal learning algorithms for playing repeated games against an adversary with unknown payoffs. In this problem, the first player (the learner) commits to a learning algorithm against a second player (the optimizer), and the optimizer best-responds by choosing the optimal-dynamic strategy for their payoff. Classic learning algorithms provide some counterfactual guarantees for the learner, but might perform much more poorly than other learning algorithms against particular optimizer payoffs. In this paper, we introduce the notion of asymptotically Pareto-optimal learning algorithms. Intuitively, if a learning algorithm is Pareto-optimal, then there is no other algorithm which performs asymptotically at least as well against all optimizers and performs strictly better (by at least $\Omega(T)$) against some optimizer. We show that well-known no-regret algorithms such as Multiplicative Weights and Follow The Regularized Leader are Pareto-dominated. However, while no-regret is not enough to ensure Pareto-optimality, we show that a strictly stronger property, no-swap-regret, is a sufficient condition for Pareto-optimality. Proving these results requires us to address various technical-challenges specific to repeated play, including the fact that there is no simple characterization ofhow optimizers who are rational in the long-term best-respond against a learning algorithm over multiple rounds of play. To address this, we introduce the idea of the asymptotic menu of learning algorithm: the convex closure of all correlated distributions over strategy profiles that are asymptotically implementable by an adversary. Interestingly, we show that all no-swap-regret algorithms share the same asymptotic menu, implying that all no-swap-regret algorithms are ``strategically-equivalent''.

3 - Efficient Prior-Free Mechanisms for No-Regret Agents

Natalie Collina, University of Pennsylvania, Philadelphia, PA, United States

We study a repeated Principal Agent problem between a long lived Principal and Agent pair in a prior free setting. In our setting, the sequence of realized states of nature may be adversarially chosen, the Agent is non-myopic, and the Principal aims for a strong form of policy regret. Following Camara, Hartline, and Johnson, we model the Agent's long-run behavior with behavioral assumptions that relax the common prior assumption (for example, that the Agent has no swap regret). Within this framework, we revisit the mechanism proposed by Camara et al., which informally uses calibrated forecasts of the unknown states of nature in place of a common prior. We give two main improvements. First, we give a mechanism that has an exponentially improved dependence (in terms of both running time and regret bounds) on the number of distinct states of nature. To do this, we show that our mechanism does not require truly calibrated forecasts, but rather forecasts that are unbiased subject to only a polynomially sized collection of events -- which can be produced with polynomial overhead. Second, in several important special cases -- including the focal linear contracting setting -- we show howto remove strong ``Alignment'' assumptions by specifically deploying ``stable'' policies that do not have any near ties that are payoff relevant to the Principal. Taken together, our new mechanism makes the compelling framework proposed by Camara et al. much more powerful, now able to be realized over polynomially sized state spaces, and while requiring only mild assumptions on Agent behavior.

4 - Algorithmic Precision and Human Decision: A Study of Interactive Optimization for School Schedules

Arthur Delarue, Georgia Institute of Technology, Atlanta, GA, United States

In collaboration with the San Francisco Unified School District (SFUSD), this paper introduces an interactive optimization framework to tackle complex school scheduling challenges. The choice of school start and end times is an optimization challenge, as schedules influence the district's transportation system, and limiting the associated costs is a computationally difficult combinatorial problem. However, it is also a policy challenge, as transportation costs are far from the only consequence of school schedule changes. Policymakers need time and knowledge to balance these considerations and reach a consensus carefully; past implementations have failed because of policy issues despite state-of-the-art optimization approaches. We first motivate our approach with a micro-foundation model of the interplay between policymakers and researchers, arguing that limiting their dependency is key. Building on these insights, we propose a framework that includes (1) a fast algorithm capable of solving the school schedule problem that compares favorably to the literature and (2) an interactive optimization approach that leverages this speed to allow policymakers to explore a variety of solutions in a transparent and efficient way, facilitating the policy decisionmaking process. The framework led to the first optimization-driven school start time changes in the US, updating the schedule of all 133 schools in SFUSD in 2021 with annual transportation savings exceeding $5 million. A comprehensive survey of approximately 27,000 parents and staff in 2022 provides evidence of the approach's effectiveness.

SA10

Summit - 330

Online Markets and Platforms

Invited Session

Auctions and Market Design

Chair: Rene Caldentey, The University of Chicago, Chicago

Co-Chair: Farbod Ekbatani, Chicago Booth School of Business, Chicago, IL, United States

1 - A Stochastic Growth Model for Online Platforms

Farbod Ekbatani, The University of Chicago, Chicago, IL, United States, Rene Caldentey

We propose a growth model of platforms operating as market markers in two-sided matching markets. The growth of these platforms, often dubbed the “chicken and egg dilemma” in the literature, poses an inherent challenge; service providers find value in joining the platform due to the presence of customers, while customers are drawn to the platform by the availability of diverse services. This endogenous network effect influences the growth trajectory and propels emerging platforms towards “get big fast” strategies to rapidly achieve critical mass on both sides. To address the limitation in the literature, we propose a dynamic model that explicitly accounts for the stochastic dynamics by which a two-sided platform gradually grows over time. We model the platform as a Markovian system, with its state space defined by s_t, the number of active service providers at time t. Service providers join the platform according to a Poisson process with an intensity $\mu(s_t)$, which is indirectly influenced by the compensation offered by the platform. Customers, on the other hand, are transitory and arrive following a Poisson process with state-dependent intensity $\lambda(s_t)$ that captures consumers' preferences for larger platforms. Under some mild conditions, we show that there are upper bounds on how fast and large the platform can grow. We also show that simple compensation rules, such as uniform payments, fail to achieve this optimal expansion. Instead, an optimal compensation mechanism must take into account the ``{\em seniority}'' of each individual active server, as measured by the time they joined the platform.

2 - Signaling Competition In Two-Sided Markets

Yuri Fonseca, Columbia University, New York, NY, United States, Omar Besbes, Ilan Lobel, Fanyin Zheng

We consider decentralized platforms facilitating many-to-many matches between two sides of a marketplace. In the absence of direct matching, inefficiency in market outcomes can easily arise. For instance, popular supply agents may garner many units from the demand side, while other supply units may not receive any match. A central question for the platform is how to manage congestion and improve market outcomes. We study the impact of a detail-free lever: the disclosure of information to agents on current competition levels. Disclosing competition reduces the perceived value of popular units, but, at the same time, it can help agents on the other side better elect across options. How large are such effects, and how do they affect overall market outcomes? We answer this question empirically. We partner with the largest service marketplace in Latin America, which sells non-exclusive labor market leads to workers. We propose a structural model which allows workers to internalize competition at the lead level and captures the equilibrium effect of such reaction to competition at the platform level. We estimate the model by leveraging agents' exogenous arrival times and a change in the platform's pricing policy. Using the estimated model, we conduct counterfactual analyses to study the impact of signaling competition on workers' lead purchasing decisions, the platform's revenue, and the expected number of matches. We find that signaling competition is a powerful lever for the platform to reduce congestion, redirecting demand, and ultimately improving the expected number of matches for the markets we analyze.

3 - Feature-Based Dynamic Matching

Akshit Kumar, Columbia Business School, New York, NY, United States, Yilun Chen, Yash Kanoria, Wenxin Zhang

Motivated by matching platforms that match agents in a centralized manner, we study dynamic two-sided matching in a setting where both customers (demand) and service providers (supply) are heterogeneous and the pool of service providers is limited. We model heterogeneity on the two sides of the market by demand weight vectors drawn i.i.d. from some distribution, and supply feature vectors drawn i.i.d. from a (possibly) different distribution. The matching of a demand-supply pair generates a matching value that depends on their weight and feature vectors. We adopt a notion of regret, specifically the additive loss relative to the value (per match) achievable in the continuum limit, as our performance metric for matching policies. Simple myopic policies suffer non-vanishing regret in large markets. We propose a forward-looking supply-aware policy dubbed Simulate-Optimize-Assign-Repeat (SOAR) which balances between producing high match value for the current match and preserving valuable supply for future customers. We prove that SOAR achieves the optimal regret scaling under different assumptions on the demand and supply distributions. En-route to proving our guarantees we develop a novel framework for analyzing the performance of our SOAR policy which may be of broader interest. As a corollary of our techniques, we also resolve an open problem posed in Kanoria 2022.

4 - Optimal Design of Recommended Donations

Francisco Castro, UCLA Anderson School of Management, Los Angeles, CA, United States, Scott Rodilitz

Nonprofit fundraising websites often display a set of recommended donation amounts, allowing prospective donors to effortlessly select an amount from this menu of recommendations instead of manually inputting their ideal donation. Although this strategy is effective at shaping behavior, it can also backfire: recommended amounts attract donors with both lower and higher ideal donations, potentially leading to a net decrease in revenue. To address this challenge, we present a comprehensive framework for designing a menu of recommendations to maximize fundraising revenue in the presence of heterogeneous donors. Our analysis reveals the limitations of a greedy approach. Instead, we design an algorithm based on dynamic programming principles that efficiently finds an optimal menu. Additionally, we shed light on the value of information by comparing against a benchmark that knows the largest amount that each donor would select. If the nonprofit has information about each donor’s ideal donation, it can obtain a constant-factor guarantee with respect to this full-information benchmark. If the nonprofit only has distributional information, we characterize how the guarantee depends on donor heterogeneity and the size of the menu. Our results highlight the value of market segmentation as well as the potential advantages of larger menus. As a case study, we apply our end-to-end methodology to experimental data from Altmann et al 2019. Our counterfactual analysis suggests that the optimal menu would increase revenue by more than 3%.

SA11

Summit - 331

Statistical Network Analysis and Applications

Invited Session

Telecommunications and Network Analytics

Chair: Jin-Zhu Yu, UT Arlington, Arlington, TX, United States

1 - Forest Expression of Networks

Yipeng Wang, University of Florida, 32608, FL, United States, Peihua Qiu

Complex networks, showing hierarchical relationships and the sparse structures between nodes, are ubiquitous in social and physical science. A common question in many studies is to visualize or analyze the network hierarchy. Currently, this is addressed by limited analytical tools, usually using ego networks or designed features to depict the network structures. For a network, we present a novel method to express connected components as trees. Thus, the network can be viewed as the forest. The 3D visualization and features based on a forest expression provide insights into network structures. Furthermore, both the 3D visualization and extracted features are compared to identify structural changes between two networks. For a dynamic network, we achieve the scalable overview by using pixel-based visualizations of the chosen features. In various cases considered, numerical studies show that proposed graphical and analytical tools are effective and flexible for network analysis.

2 - Network Clustering and Machine Learning for Predicting Bacterial Protein Functions

Mishkatur Rahman , North Dakota State University, Fargo, ND, United States, Harun Pirim, Hasan Tekedar, Larry Hanson, Matt Griffin

Understanding the complex network of protein interactions is key to uncovering cellular functions and genetic foundations. This study introduces a thorough approach that includes bacterial protein network clustering leveraging functional similarity and a predictive machine learning model to identify the functions of bacterial proteins. The methodology centers around developing mixed integer linear programming (MILP) clustering model to reveal groups of proteins that interact with each other by maximizing their functional similarity within the cluster. The parameters of the MILP model are predicted using node embeddings from protein-protein interaction network for a bacterial protein generated from STRING database. Validation through gene sequence alignment further demonstrated the effectiveness of the functional similarity-based MILP clustering method, revealing a notable increase in the homogeneity of gene sequences within clusters. In parallel, a predictive model is developed to determine the molecular function of bacterial proteins. This model extracts features from protein sequences, annotations, motifs, and physicochemical properties to accurately predict associated GO terms. By employing advanced machine learning techniques, the model optimizes the accuracy of protein function predictions for newly characterized bacterial proteins. The integration of network clustering and predictive modeling creates a strong foundation for investigating gene functionalities and predicting protein functions in bacterial systems, thus advancing the field of bioinformatics and providing valuable insights for genetic research and molecular biology. This study highlights the importance of computational methods in understanding the complex relationships and functionalities within genetic networks.

3 - Machine Learning and Network Analysis to Predict Hypothetical Protein Functions of Aeromonas hydrophila

Zaidur Rahman, University of Arkansas, Fayetteville, AR, United States

Aeromonas hydrophila, antibiotic resistant gram negative bacteria, is a major fish pathogen. Moreover, A. hydrophila is considered to cause 13% of gastroenteritis cases in the United States. Therefore, it is important to identify groups of proteins that are effective in antibiotic resistance and causing mortality in aquaculture. We train machine learning models on existing A. hydrophila genomes to predict functions of 83 carefully filtered hypothetical proteins. Network analysis is conducted to cluster these proteins based on their similarities. Both ML and network analysis inform about possible roles of these proteins in antibiotic resistance and fish mortality.

SA12

Summit - 332

Supply Chain and Inventory Optimization

Flash Session

Contributed

Chair: xiaona zhang, Harbin Engineering University, Harbin, N/A

1 - Control and Enforcement in a Public Sector Direct Delivery Supply Chain, Considering Horizontal Interaction and Supply Disruption

Matan Shnaiderman, Bar-Ilan University, Ramat-Gan, Israel, Liron Ben-Baruch

Our research deals with Direct Vendor Delivery Supply Chains - common in public and government sector organizations and agencies, such as hospitals, universities and public offices. We consider multiple retailers/facilities and the possibility of limited supply available to a supplier, and we consider behavioral operations factors affecting a supply chain member's performance and decisions. The retailers cannot directly enact sanctions on the supplier and need to persuade the purchasing department's representative to penalize the supplier, due to contract breaches, by investing time in enforcement efforts. We review different scenarios, featuring supply disruptions, different inventory allocation methods, and different attentiveness levels of the organization's purchasing department representative to the retailers' enforcement efforts against the supplier. When the supplier has a limited amount of inventory available for retailers, informing the retailers about the expected shortage and the chosen inventory allocation method enables them to prepare for the shortage and set the optimal values in their decision parameters. Competition between the retailers may create horizontal competition in the supply chain, which benefits the supplier but creates excess costs for the retailers. On the other hand, cooperation between retailers allows reducing inventory costs and lowering the risk of shortages among retailers. The collaboration includes determining a centralized strategy for the level of each retailer's safety stock order, and a plan for stock sharing between the retailers. Cooperation between retailers lowers the risk of shortages and improves the profitability of the retailers, usually at the expense of the supplier.

2 - Incentives on Green Supply Chain considering Product Design and Channel Encroachment

lin li, Beijing Jiaotong University, Beijing, China, People's Republic of

We consider the problem of determining the online channel encroachment decision and the product's greenness level of the manufacturer when the retailer can adopt different product incentive contracts. Both for the encroachment and non-encroachment setting, we formulate a green dual-channel supply chain composed of a manufacturer and a retailer. We find that for manufacturers, when the online channel's invasion cost is small and the green technology cost coefficient is large, or when the invasion cost is large and the green technology cost coefficient is moderate, the retailer adopting green reinvestment strategy and the manufacturer entering the online channel dominates. When the invasion cost is large or the green technology cost coefficient is small, the retailer adopting green reinvestment strategy and the manufacturer no entering the online channel dominates. When the invasion cost is moderate and the green technology cost coefficient is large, the retailer adopting wholesale price strategy or cost-sharing strategy and the manufacturer entering the online channel dominates. When both the invasion cost and the cost coefficient are small, the retailer adopting wholesale price strategy and the manufacturer entering the online channel dominates. For the retailer, if the manufacturer decides not to enter the online channel and the green technology cost coefficient is small or large, the retailer benefits the most from green reinvestment strategy. Conversely, the retailer benefits the most from cost-sharing strategy.

3 - Distributionally Robust Optimization for Overseas Warehouse Location Problem under Decision-Dependent Demand Uncertainty

xiaona zhang, Harbin Engineering University, Harbin, China, People's Republic of

In global supply chain management, the decision-making process for the location of overseas warehouses plays a crucial role in enhancing logistical efficiency, optimizing costs, and improving customer satisfaction. However, the uncertainty of market demand and its interaction with the decision-making process for warehouse location present challenges to the site selection issue.This paper constructs a decision-dependent demand ambiguity set based on moment information, interpreting the demand moments as a function of the location decision, and proposes a distributionally robust optimization model under decision-dependent demand uncertainty.To solve this model, the paper employs robust counterpart theory, transforming the complex distributionally robust optimization problem into a mixed-integer linear programming problem, thereby simplifying the solution process and enhancing solution efficiency.Through extensive computational studies, we have found that the method proposed in this paper exhibits outstanding performance in uncertain environments. Not only can it enhance the overall benefits and service quality of the supply chain, but it also provides decision support for the rational layout of overseas warehouses under uncertain conditions.

SA13

Summit - 333

Intelligent Transportation Systems

Flash Session

Flash

Chair: Tristan Ford

1 - Real-time driver behavior monitoring using the integration of transfer learning and Extended Reality

Amirarash Kashef, Mississippi State University, Starkville, MS, United States, Junfeng Ma

According to the World Health Organization (WHO) about 1.19 million individuals lose their lives annually due to road traffic accidents, while between 20 and 50 million others sustain non-lethal injuries, often resulting in disabilities. In fact, one of the main reasons is distracted drivers and the distraction caused by mobile phones is a growing concern for road safety. In this study, we have made use of the image-based distracted driver dataset V2 with 44 different drivers in 10 different classes. The data used is in RGB format and belongs only from the side-view camera that captured the driver’s body. Additionally, we have extended the dataset by adding a class for drowsy drivers which statistics show an estimated 17.6 percent of fatal car crashes between 2017 and 2021 involve a drowsy driver. Transfer learning was then implemented on the dataset for classification purposes. Our proposed Deep Convolutional Neural Network (DCNN) model which is based on MobileNet has shown %93.60 accuracy and 0.94 F1-score. Subsequently, the trained AI model is then installed on Extended Reality (XR) device through a series of procedures to enable real-time driver behavior monitoring.

2 - Go Smart and Green: A Sustainable Traffic Management Framework Utilizing Vehicle Energy Consumption and Emission Patterns

Sudenaz Ozvural, Drexel University, Philadelphia, PA, United States, Liang Zhang, Shiyi Zhang

The increasing number of vehicles on roads has reduced mobility, increased energy consumption, and emphasized environmental concerns. Especially in congested urban areas, vehicles tend to consume more gasoline and emit higher levels of pollutants. In this research, we introduce a traffic signal control system that utilizes vehicle energy consumption and emission data. Our primary objective is effectively balance and reduce vehicle energy consumption and emissions at signalized intersections while ensuring equitable allocation of resources (such as green time and "travel credits") among different vehicle types. Specifically, the system collects vehicle energy consumption and emission data through sensors installed on vehicles, which transfer this information to the control center as the input. The control center outputs signal timing plans based on resource allocation algorithms. Additionally, fairness indices are established for various vehicle types (including trucks, regular vehicles, and electric vehicles). Each vehicle is allocated "travel credits" for future use. Simulation experiments have demonstrated that our proposed method can significantly reduce both energy consumption and emissions simultaneously.

3 - Preventing Pedestrian Road Fatalities on Indian Reservations

Bukola Bakare, Middle Tennessee State University, Murfreesboro, TN, United States, Chris Bic Byaruhanga

Due to cultural, economic, and historic factors, American Indians tend to walk to their destinations more frequently than other populations in the United States. They also have a very high rate of pedestrian deaths due to crashes involving motor vehicles. Contributing factors in these crashes include limited sight zones, alcohol consumption, lack of sidewalks, and uncontrolled intersections. This research investigates crashes on American Indian lands and in other rural locations to provide insight into how to curtail future pedestrian injuries in tribal communities. We utilized National Highway Traffic Safety Administration (NHTSA) and the Bureau of Indian Affairs (BIA) data from the years 2016 to 2020. Descriptive analysis of crash records is used in a diagnostic process to distinguish elements associated with motor vehicle crashes involving pedestrians on tribal lands and those who are indigenous. The pedestrian-involved crashes will be compared to similar incidents on other roads, to distinguish significant differences in the driver involved, vehicle, environment, and roadway elements. Gaining a better understanding of the circumstances surrounding pedestrian/vehicle crashes on American Indian lands will help decision makers to select the most appropriate, evidence-based solutions to these incidents.

4 - Location-Allocation of Emergency Service Systems in Light Traffic: Application to Lift-Trap Rescue

Xin Wang, National University of Singapore, Singapore, Singapore, Weiliang Liu, Zhisheng Ye

This work addresses location-allocation challenges in emergency service systems (ESSs) under light traffic conditions. Typically, assessing ESS performance relies on response time metrics, complicated by the distribution of queued calls. We tackle this complexity by analyzing spatial queues as arrival rates approach zero, yielding simple performance expressions. Leveraging these insights, we formulate tractable optimization models for ESS location-allocation, demonstrating significant improvements over benchmark approaches. Real-world application in Singapore's lift-trap rescue highlights a circa 60% reduction in annual penalized incidents compared to the status quo.

5 - A Robust Time-dependent Green Location-Routing Problem with Time Windows

Seyedehsaba Siadati, Eindhoven University of Technology, Eindhoven, Netherlands, Virginie Lurkin, Mehrdad Mohammadi, Tom Van Woensel

This study addresses the dynamic challenges of urban transportation in logistics caused by time-varying traffic congestion, introducing the Robust Time-Dependent Green Location-Routing Problem with Time Windows (R-TDGLRP-TW). A key innovation is incorporating a time-dependent speed function, strategically capturing the nuanced dynamics of urban transportation. Additionally, the inherent uncertainty in travel speed (time), which significantly impacts the system'sperformance, is considered in a robust optimization context.

The R-TDGLRP-TW optimizes hubs, vehicle routes, and delivery schedules, minimizing economic and environmental costs. To tackle this, we develop a mixed-integer linear programming model for R-TDGLRP-TW, incorporating the linearization of dynamic programming recursive equations proposed in the literature into a deterministic formulation. A matheuristic solution algorithm is then used to solve large-scale instances efficiently within a reasonable computation time. Our focus on the dynamic aspects of urban transportation aims to efficiently provide practical insights, helping service providers navigate the complex trade-offs encountered in logistics planning.

6 - Delay management in transit vehicle scheduling

Tristan Ford, University of British Columbia, Vancouver, BC, Canada, Julia Yan, Amy Kim

Disruptions from congestion, construction, and weather are a major source of primary delay in public transit networks. These delays may then propagate along a vehicle's schedule if there is not sufficient buffer time to absorb delay. In this work, we model a vehicle scheduling problem with propagated delay, using stochastic optimization to minimize operating costs while being robust to propagated delay. We analyze a small network in British Columbia, Canada to highlight best practices and guidelines for transit agencies in managing delay.

SA14

Summit - 334

Financial Engineering and FinTech

Invited Session

Finance

Chair: Steven Kou, Boston University, Boston, MA, 02215, United States

1 - Equilibrium Dividend and Capital Policy Under TIME-Inconsistent Preferences

SANG HU, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of

This study investigates the dividend payment and capital injection problem under time-inconsistent preferences. The objective of shareholders is to maximize expected total discounted dividends, while the capital injection with cost is to reduce the likelihood of the insurance company being bankruptcy. The problem is time-inconsistent with non-exponential discounting and we employ the definition of weak equilibrium in stochastic control to be the equilibrium strategy. The extended system of HJB equations are established for general non-exponential discounting. We also consider the pseudo-exponential discounting as an example and derive the analytical treatment.

2 - Inverse Leverage Effect for Cryptocurrencies and Meme Stocks: A Comprehensive Framework

Steven Kou, Boston University, Boston, MA, United States

Although the leverage effect, i.e., a negative correlation between the return and volatility, and the inverse leverage effect have been suggested for equities and commodities, respectively, the existing studies suffer from an endogeneity problem because they only model one asset. By using a comprehensive multivariate model with jumps and heavy tail distribution for both an equity index and the asset, we find inverse and threshold inverse leverage effects for cryptocurrencies and meme stocks. Network effects cannot explain this finding. To handle over 18,000 latent variables in the model, a particle Gibbs with an ancestor sampling algorithm is extended to estimate parameters efficiently.

3 - Price Discovery on Decentralized Exchanges

Agostino Capponi, Columbia University, New York, NY, United States, RUIZHE JIA, Shihao Yu

How would informed traders trade if they have to publicly compete for execution? Decentralized exchanges (DEXs) provide an ideal laboratory to answer this question, as they require traders to publicly submit fees to prioritize their orders. We provide empirical evidence that informed traders do not bid low fees to hide their trading intentions. Rather, they publicly bid high fees to signal their information, even if private bidding is an option. Using a unique dataset of Ethereum mempool orders, we show that informed traders do so by employing a “jump bidding” strategy, where they place high initial bids to deter potential competitors.

4 - Asset Pricing with A-Maxmim Expected Utility Model

Xuedong He, The Chinese University of Hong Kong, Shatin, Hong Kong, Jiacheng Fan, Ruocheng Wu

We study an asset pricing problem in which a representative agent trades a risky stock, a risk-free asset, and human capital to maximize her preference value of consumption represented by the α-maxmin expected utility model. This preference model is known to lead to time inconsistency, so we consider intra-personal equilibrium for the representative agent and define the market equilibrium to the set of asset prices under which the intra-personal equilibrium strategy clears the market. We prove that there exists a unique market equilibrium and the asset prices are determined by the solution to a second-order ordinary differential equation. Finally, we conduct comparative statics to study the effect of the agent's ambiguity attitude on the asset prices.

SA15

Summit - 335

Deep Reinforcement Learning for Solving Inventory and Supply Chain Problems

Invited Session

Revenue Management and Pricing

Chair: Linwei Xin, Booth School of Business, University of Chicago, Chicago, IL, United States

1 - Vc Theory for Inventory Policies

Yaqi Xie, University of Chicago, Chicago, IL, United States, Will Ma, Linwei Xin

Advances in computational power and AI have increased interest in reinforcement learning approaches to inventory management. This paper provides a theoretical foundation for these approaches and investigates the benefits of restricting to policy structures that are well-established by decades of inventory theory. We prove generalization guarantees for learning several well-known classes of inventory policies, including base-stock and (s, S) policies, by leveraging the celebrated Vapnik-Chervonenkis (VC) theory. We apply the concepts of the Pseudo-dimension and Fat-shattering dimension from VC theory to determine the generalizability of inventory policies. We focus on a classical setting without contexts, but allow for an arbitrary distribution over demand sequences and do not make any assumptions such as independence over time. We corroborate our supervised learning results using numerical simulations.

Managerially, our theory and simulations translate to the following insights. First, there is a principle of “learning less is more” in inventory management: depending on the amount of data available, it may be beneficial to restrict oneself to a simpler, albeit suboptimal, class of inventory policies to minimize overfitting errors. Second, the number of parameters in a policy class may not be the correct measure of overfitting error: the class of policies defined by T time-varying base-stock levels exhibits a generalization error comparable to that of the two-parameter (s, S) policy class. Finally, our research suggests situations where it could be beneficial to incorporate the concepts of base-stock and inventory position into black-box learning machines, instead of having these machines directly learn the order quantity actions.

2 - Learning An Inventory Control Policy with General Inventory Arrival Dynamics

Sohrab Andaz, Amazon, New York, NY, United States, Carson Eisenach, Dhruv Madeka, Kari Torkkola, Randy Jia, Dean Foster, Sham Kakade

In this work, we address the problem of learning and backtesting inventory control policies in the presence of general arrival dynamics, which we term as a quantity-over-time arrivals model (QOT). We also allow for order quantities to be modified as a post-processing step to meet vendor constraints such as order minimum and batch size constraints, a common practice in real supply chains. Building upon recent work, we similarly formulate the periodic review inventory control problem as an exogenous decision process, where most of the state is outside the control of the agent. Madeka et al. (2022) show how one can construct a simulator that replays historic data to solve problems in this setting via deep reinforcement learning. In our case, we incorporate a deep generative model for the arrivals process as part of the history replay. By formulating the problem as an exogenous decision process, we can apply results from Madeka et al. (2022) to obtain a reduction to supervised learning. Via simulation studies, we show that this approach yields statistically significant improvements in profitability over production baselines. Using data from a real-world A/B test, we show that Gen-QOT generalizes well to off-policy data and that the resulting buying policy outperforms traditional inventory management systems in real-world settings.

3 - Neural Inventory Control in Networks via Hindsight Differentiable Policy Optimization

Matias Alvo, Columbia University, New York, NY, United States, Daniel Russo, Yash Kanoria

We argue that inventory management presents unique opportunities for reliably applying and evaluating deep reinforcement learning (DRL). Toward reliable application, we emphasize and test two techniques. The first is Hindsight Differentiable Policy Optimization (HDPO), which performs stochastic gradient descent to optimize policy performance while avoiding the need to repeatedly deploy randomized policies in the environment—as is common with generic policy gradient methods. Our second technique involves aligning policy (neural) network architectures with the structure of the inventory network. Specifically, we focus on a network with a single warehouse that consolidates inventory from external suppliers, holds it, and then distributes it to many stores as needed. In this setting, we introduce the symmetry-aware policy network architecture. We motivate this architecture by establishing an asymptotic performance guarantee and empirically demonstrate its ability to reduce the amount of data needed to uncover strong policies. Toward rigorous evaluation, we create and share new benchmark problems, divided into two categories. One type focuses on problems with hidden structures that allow us to compute or bound the cost of the true optimal policy. Across four problems of this type, we find HDPO consistently attains near-optimal performance, handling up to 60-dimensional raw state vectors effectively. The other type of evaluation involves constructing a test problem using real time series data from a large retailer, where the optimum is poorly understood. Here, we find HDPO methods meaningfully outperform a variety of generalized newsvendor heuristics. Our code can be found at https://github.com/MatiasAlvo/Neural_inventory_control.

4 - Massive Speedups for Policy Simulation in Supply Chain Reinforcement Learning

Aryan Khojandi, Massachusetts Institute of Technology, Cambridge, MA, United States, Joren Gijsbrechts, Vivek Farias, Tianyi Peng, Andrew Zheng

Simulating a single trajectory of a dynamical system under some state-dependent policy is a core bottleneck in policy optimization (PO) algorithms. The many inherently serial policy evaluations that must be performed in a single simulation constitute the bulk of this bottleneck. To wit, in applying PO to supply chain optimization (SCO) problems, e.g. in inventory-management settings, simulating a single month of a supply chain can take several hours. We present an iterative algorithm for policy simulation, which we dub Picard Iteration. This scheme carefully assigns policy-evaluation tasks to independent processes. Within an iteration a single process evaluates the policy only on its assigned tasks while assuming a certain 'cached' evaluation for other tasks; the cache is updated at the end of the iteration. Implemented on GPUs, this scheme admits batched evaluation of the policy on a single trajectory. We prove that the structure afforded by many SCO problems allows convergence in a small number of iterations *independent* of the horizon. We demonstrate practical speedups of 400x on large-scale SCO problems using a single GPU, and also demonstrate practical efficacy in other RL environments.

SA16

Summit - 336

Efficient Learning Methods and Revenue Management Applications

Invited Session

Revenue Management and Pricing

Chair: Heng Zhang, Arizona State University, Tempe, AZ, United States

Co-Chair: Mengxin Wang, The University of Texas at Dallas, Richardson, TX, United States

1 - Deep-Learning-Based Causal Inference for Large-Scale Combinatorial Experiments: Theory and Empirical Evidence

Heng Zhang, Arizona State University, Tempe, AZ, United States, Zikun Ye, Zhiqi Zhang, Dennis J. Zhang, Renyu Zhang

Large-scale online platforms conduct numerous A/B tests daily, but not all treatment combinations are tested, posing challenges in estimating causal effects and identifying optimal combinations. We introduce debiased deep learning (DeDL), a framework that combines deep learning with doubly robust estimation to estimate the causal effects of unobserved treatment combinations using limited data. DeDL provides efficient, consistent, and asymptotically normal estimators. We validated the framework through collaboration with a video-sharing platform. DeDL outperformed benchmarks in estimating ATEs and identifying the optimal combination.

2 - Decoupling Learning and Decision-Making: Breaking the O(\sqrt{T}) Barrier in Online Resource Allocation with First-Order Methods

Chunlin Sun, Stanford University, Stanford, CA, United States, Wenzhi Gao, Chenyu Xue, Dongdong Ge, Yinyu Ye

Online linear programming plays an important role in both revenue management and resource allocation, and recent research has focused on developing efficient first-order online learning algorithms. Despite the empirical success of first-order methods, they typically achieve a regret no better than $\mathcal{O}(\sqrt{T})$, which is suboptimal compared to the $\mathcal{O}(\log T)$ bound guaranteed by the state-of-the-art linear programming (LP)-based online algorithms. This paper establishes several important facts about online linear programming, which unveils the challenges for first-order-method-based online algorithms to achieve beyond $\mathcal{O}(\sqrt{T})$ regret. To address these challenges, we introduce a new algorithmic framework that decouples learning from decision-making. For the first time, we show that first-order methods can attain regret $\mathcal{O}(T^{1/3})$ by this new framework. Lastly, we conduct numerical experiments to validate our theoretical findings.

3 - Workforce scheduling and pricing problem in crowdsourced delivery under uncertainty

MINGYAO QI, Tsinghua University, Shenzhen International Graduate School, Shenzhen, China, People's Republic of, Haoyue Liu, Sheng Liu, Zuo-Jun Max Shen

Crowdsourced labor has significantly changed the on-demand delivery market, as gig workers can supplement in-house couriers to fulfill orders. Motivated by the high level of uncertainty in on-demand instant delivery, we model and solve a joint optimization on workforce scheduling and pricing problem (WSPP) with three types of labor, namely in-house couriers, scheduled couriers, and ad hoc couriers for last-mile delivery system. We propose an integrated stochastic programming, robust optimization, and learning model on scheduled couriers' employment, order assignment, and compensation pricing for ad hoc couriers while considering three types of uncertainties, including order patterns, the location of ad hoc couriers, and travel time. We devise an improved logic-based benders decomposition nested column-and-constraint generation algorithm based on stepwise tightening relaxation to solve WSPP exactly, accelerated by a series of valid inequalities and strategies. We also design various approximation algorithms encompassing shift separating, rolling horizon, and rolling horizon with lookahead, which deliver high-quality solutions in a shorter amount of time. As indicated by the numerical experiments, our approach outperforms the solver and provides optimal solutions within the time limit for most instances. The proposed integrated model has more robustness and lower cost compared to deterministic model when faced with uncertainty. The results on real-world datasets show that the uncertainty of ad hoc courier locations significantly impacts decisions for the number of scheduled couriers, leading to several managerial insights.

4 - Model Stacking with Flexible Data Aggregation: Inventory Management in Meituan’s Flash Sales Mode

Zhenkang Peng, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Chengzhang Li, Ying Rong, Zichao Luo, Mingyong Zhao, Guangrui Ma, You Li

Given the rapid pace of product launches and shortened product cycles, although the company accumulates a substantial volume of data overall, the data available per product remains limited on average. Consequently, the task of decision-making across thousands of products, each with scant historical data, has emerged as an urgent and new challenge. To tackle this challenge, we have devised a strategy known as Model Stacking with Flexible Data Aggregation (MOST-FLEXDA). This approach includes data aggregation from multiple perspectives, coupled with the selection of specific models customized for these distinct partitions. Subsequently, we utilize a linear model to stack the outputs from these diverse models. Finally, our methodology is put to the test on the Meituan Youxuan platform, which primarily evaluates performance based on two metrics: the first being the ratio of potential profit loss to maximum profit (LSR), and the second being the ratio of remaining inventory to total demand (IR). The results demonstrate that, compared to the existing algorithms utilized by Meituan, our proposed method achieves an average reduction of 42% in IR and a 14% decrease in LSR for newly launched products. Additionally, for standard new products, it also yields reductions of approximately 10% in both IR and LSR.

5 - Bayesian dynamic pricing and the optimality of incomplete learning

Meixi Wu, National University of Singapore, Singapore, Singapore, Yifan Feng, Zhenyu Hu

In this paper, we introduce a dynamic pricing model with demand learning. We consider a setting where a monopolist makes pricing decisions facing a stream of customers over an infinite horizon. The customers are short-lived and homogenous, all of whom share the same private valuation for the product. Initiating with a prior on the valuation, the monopolist updates her belief through observation of customers' purchase decisions. By formulating this problem as a Bayesian dynamic programming problem, we characterize the optimal pricing policy that notably identifies an absorbing region of beliefs. Within this region, it becomes optimal for the monopolist to cease further learning and adopt a static, myopic pricing strategy. This cessation hinges on the discount factor being less than one, and, under a uniform prior, there is a positive probability of transitioning into this no-learning region.

SA17

Summit - 337

Emerging Topics in Revenue Management: Digital Innovation, Estimation and Optimization

Invited Session

Revenue Management and Pricing

Chair: Guang Li, Queen’s University, Kingston, ON, Canada

Co-Chair: Elaheh Fata, Queen's University, Kingston, ON, Canada

1 - Demand Estimation with Product Bundles

Xianfeng Meng, Smith School of Business, Queen's University, Kingston, ON, Canada, Sumit Kunnumkal, Murray Lei, Guang Li, Anton Ovchinnikov

Traditionally, firms estimate product demand elasticity by varying prices and observing demand changes. However, this method poses challenges for firms such as chain stores and restaurants, where price alterations can be impractical. This study introduces a novel method under the multinomial logit (MNL) consumer choice model, where firms can infer the demand elasticity of new products through discounted product bundles without adjusting individual item prices. We analytically demonstrate that this method can sometimes yield more accurate estimations than direct item discounts. Through numerical analysis with synthetic data, we investigate the accuracy of direct discount estimators versus bundle discount estimators, examining their performance across various product choice set (menu) sizes and their robustness to differences between the firm’s model and actual consumer behavior. An upcoming empirical study in collaboration with a local restaurant will further explore the practical application of bundle discounts in estimating the real-world demand elasticity.

2 - Content-Centered OR Creator-Centered? An Economic Analysis of Fan Loyalty on User-Generated Content Platforms

Meilin Gu, Tsinghua University, Beijing, China, People's Republic of, Zizheng Liu, Dengpan Liu

The loyalty of fan communities to content creators significantly impacts stakeholders on user-generated content (UGC) platforms. Notably, major UGC platforms exhibit varying stances on creators’ cultivation of fan loyalty. For instance, Douyin (the Chinese version of TikTok) hinders creators from building robust fan followings, while Kuaishou (another major UGC platform in China) takes active measures to encourage and support creators in cultivating fan loyalty on its platform. This stark contrast has sparked an ongoing debate about the role of UGC platforms in the cultivation of fan loyalty. In this paper, using a game-theoretic model, we explore the optimal strategies embraced by platforms in managing creators’ cultivation of fan loyalty and their implications for key stakeholders. Our findings challenge the conventional wisdom that suggests UGC platforms should always facilitate creators in cultivating fan loyalty, as such assistance may sometimes hurt the platforms. Furthermore, our study generates meaningful implications by providing guidelines for UGC platforms to strategically manage creators’ cultivation of fan loyalty and for creators to balance their efforts between cultivating fan loyalty and enhancing content quality.

3 - Assortment Planning with N-Pack Purchasing Consumers

Ying Cao, Penn State University - Erie, Erie, PA, United States, Dorothee Honhon

For many product categories, consumers often buy multiple differentiated products on a given store visit for staggered consumption until the next store visit. In this paper, we study the assortment planning problem for a single product category when a retailer faces multi-item purchasing, so-called “n-pack” consumers as introduced by Fox et al (2018). We obtain interesting properties of the product demand functions and the optimal assortment. We also find the optimal assortment structures in four settings which correspond to four ways by which the retailer could misjudge the demand from consumers. In addition, we conduct a numerical study to investigate the differences in optimal assortment variety and calculate optimality gaps from the retailer misjudging consumer demand. Finally, we evaluate our model on a real-life data set and find that the demand proportions predicted by our model can be made extremely close to the actual proportion of sales; then we provide recommendations regarding the introduction of new flavors.

4 - Competitive Pricing in the Presence of Manipulable Information in Online Platforms

Yuqi Yang, Arizona State University, Tempe, AZ, United States, Harish Guda, Hongmin Li

To entice customers to purchase, sellers on online platforms often misrepresent the quality of their goods/services, e.g., by manipulating consumer opinion. We analyze an oligopoly where sellers, heterogeneous in their true quality, compete by jointly choosing their prices and the extent of manipulation. We solve for the unique equilibrium when price-setting firms can manipulate their perceived quality and characterize the set of sellers that manipulate in equilibrium. We identify an index called the propensity to manipulate, based on model primitives to identify the set of sellers who have greater incentive to manipulate, and show that the set of sellers that manipulate in equilibrium is upward-closed in the propensity to manipulate. The extant literature has been mixed in its findings on which sellers have greater incentive to manipulate. Our work helps reconcile the differing viewpoints in the extant literature by providing a unified perspective. We demonstrate the practical relevance of our model by mapping it to an environment consisting of sellers who are differentiated in a star-rating system based on their true rating and the volume of ratings. Depending on a seller’s rating and volume of ratings, we identify three distinct regions that arise: a cost-prohibitive region, a cost-dominant region, and a benefit-dominant region. The ability to map a seller to one of these regions allows platform managers to understand a seller’s tendency to manipulate consumer opinion dynamically over time.

SA18

Summit - 338

Learning in Sharing Economy and E-Commerce

Invited Session

Revenue Management and Pricing

Chair: Evelyn Xiao-Yue Gong, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Two-Sided Flexibility in Platforms

Jiayu (Kamessi) Zhao, MIT, Cambridge, MA, United States, Daniel Freund, Sebastien Martin

Flexibility is a cornerstone of operations management, crucial to hedge stochasticity in product demands, service requirements, and resource allocation. In two-sided platforms, flexibility is also two-sided and can be viewed as the compatibility of agents on one side with agents on the other side. Platform actions often influence the flexibility on either the demand or the supply side. But how should flexibility be jointly allocated across different sides? Whereas the literature has traditionally focused on only one side at a time, our work initiates the study of two-sided flexibility in matching platforms. We propose a parsimonious matching model in random graphs and identify the flexibility allocation that optimizes the expected size of a maximum matching. Our findings reveal that flexibility allocation is a first-order issue: for a given flexibility budget, the resulting matching size can vary greatly depending on how the budget is allocated. Moreover, even in the simple and symmetric settings we study, the quest for the optimal allocation is complicated. In particular, easy and costly mistakes can be made if the flexibility decisions on the demand and supply side are optimized independently (e.g., by two different teams in the company), rather than jointly. To guide the search for optimal flexibility allocation, we uncover two effects, flexibility cannibalization, and flexibility abundance, that govern when the optimal design places the flexibility budget only on one side or equally on both sides. In doing so we identify the study of two-sided flexibility as a significant aspect of platform efficiency.

2 - How Not to Overpackage? -- Towards a Sustainable HelloFresh Service Supply Chain.

Evelyn Xiao-Yue Gong, Carnegie Mellon University, Pittsburgh, PA, United States, Michael Johnson

Meal kit services have been hot and trending, especially among the younger generation. However, overpackaging is a common major challenge in these services. Packaging materials, including ice packs and liners, ensure the quality of the meal-kits delivered; yet too much packaging would leave a large carbon footprint and impose psychological burdens on many customers. This paper investigates artificial intelligence solutions to adaptively make the packaging decision for each box and mitigate potential overpackaging for HelloFresh, the world’s largest meal-kit company and integrated food solutions group. We design contextual bandit algorithms that take advantage of the special structures we find in the packaging problem and various contextual information such as transit conditions and box contents. Theoretically, our algorithm Contextual One-Sided Arm Elimination achieves the optimality guarantee with an $O(\sqrt{T})$ regret bound. Practically, we experiment with HelloFresh's real delivery datasets that contain hundreds of millions of records, identify and correct for issues such as confounding, and test our algorithm's performance. Given the enormous scale of HelloFresh's operations, our contextual bandit algorithm could potentially save millions of units of packaging materials per year, as well as the associated cost, energy and labor.

3 - LEGO: Optimal Online Learning under Sequential Price Competition

Shukai Li, Northwestern, Evanston, IL, United States, Cong Shi, Sanjay Mehrotra

We consider price competition among multiple sellers over a selling horizon of T periods. In each period, sellers simultaneously offer their prices and subsequently observe their respective demand that is unobservable to competitors. The realized demand of each seller depends on the prices of all sellers following a private unknown linear model. We propose a least-squares estimation then gradient optimization (LEGO) policy, which does not require sellers to communicate demand information or coordinate price experiments throughout the selling horizon. We show that our policy, when employed by all sellers, leads at a fast convergence rate $O(1/\sqrt{T})$ to the Nash equilibrium prices that sellers would reach if they were fully informed. Meanwhile, each seller achieves an optimal order-of-$\sqrt{T}$ regret relative to a dynamic benchmark policy. Our analysis further shows that the \emph{unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition} and forces regret to the order of $\sqrt{T}$ in the worst case. If each seller knows their individual price sensitivity coefficient, then a gradient optimization policy can achieve an optimal order-of-$\frac{1}{T}$ convergence rate to Nash equilibrium as well as an optimal order-of-$\log T$ regret.

4 - Queueing Matching Bandits with Preference Feedback

Jung-hun Kim, Seoul National University, Seoul, Korea, Republic of, Min-hwan Oh

In this study, we consider multi-class multi-server asymmetric queueing systems consisting of N queues on one side and K servers on the other side, where jobs randomly arrive in queues at each time. The service rate of each job-server assignment is unknown and modeled by a feature-based Multi-nomial Logit (MNL) function. At each time, a scheduler assigns jobs to servers, and each server randomly serves at most one job based on its preferences over the assigned jobs. The primary goal of the scheduler is to stabilize the queues in the system while learning the service rates of servers. For this purpose, we propose algorithms based on UCB and Thompson Sampling, which achieve system stability with a time average queue length bound of $O(N/\epsilon)$ for large enough horizon time T. Furthermore, the algorithms achieve sublinear regret bounds of $\tilde{O}(\min\{\sqrt{T}Q_{\max},T^{3/4}\})$ where $Q_{\max}$ represents the maximum queue length over agents and times. Lastly, we provide experimental results to demonstrate the performance of our algorithms.

5 - A Hybrid Sampling-Based and Gradient Descent Learning Method and Its Applications in Inventory Management

Kairen Zhang, University of Science and Technology of China, Hefei, China, People's Republic of, Zhanyue Wang, Xiangyu Gao, Sean Zhou

We propose an algorithm that marries Sample Average Approximation and Stochastic Gradient Descent for some structured online learning problems. The special structure of these problems manifests itself in important online inventory control problems where the demand distribution is unknown to the decision maker. In particular, the online control of two-echelon series systems and that of dual sourcing systems with lead time difference equal to one. For the above two problems, our algorithm yields regrets growing at the order of log T times square root of T at most where T is the length of the planning horizon. These upper bounds of regrets differ from the corresponding lower bounds by a factor of log T asymptotically. We establish these upper bounds of regrets via leveraging the additive convexity hidden in the one-period expected cost functions and constructing bridging problems. If the one period demand is greater than some positive number almost surely, then we can further improve these upper bounds and closing the gaps between upper bounds and lower bounds. In numerical experiments, our algorithm demonstrates superior performance to other algorithms in the literature.

SA19

Summit - 339

Learning in Distributed and Dynamic Environments

Invited Session

Revenue Management and Pricing

Chair: Ankur Mani, University of Minnesota - Twin Cities, Minneapolis, MN, United States

Co-Chair: Varun Gupta, Northwestern University, Evanston, IL, United States

1 - Autobidders with Budget and ROI Constraints: Efficiency, Regret, and Pacing Dynamics

Mengxiao Zhang, University of Southern California, Los Angeles, CA, United States, Brendan Lucier, Sarath Pattathil, Alex Slivkins

We study a game between autobidding algorithms that compete in an online advertising platform. Each autobidder is tasked with maximizing its advertiser’s total value over multiple rounds of a repeated auction, subject to budget and/or return-on-investment constraints. We propose a gradient-based learning algorithm that is guaranteed to satisfy all constraints and achieves vanishing individual regret. Our algorithm uses only bandit feedback and can be used with the first- or second-price auction, as well as with any "intermediate" auction format. Our main result is that when these autobidders play against each other, the resulting expected liquid welfare over all rounds is at least half of the expected optimal liquid welfare achieved by any allocation. This holds whether or not the bidding dynamics converges to an equilibrium and regardless of the correlation structure between advertiser valuations.

2 - Dynamic control under non-stationarity: Stochastic and Robust perspectives

Varun Gupta, Northwestern University, Evanston, IL, United States, Yuwei Luo, Jing Yu, Mladen Kolar, Adam Wierman

We will present two vignettes on the problem of online control of the Linear Quadratic Regulator (LQR) problem when the dynamics are non-stationary and unknown. LQR is arguably the simplest Markov Decision Process, and serves as a fertile ground for developing new frameworks for studying online and robust control policies. In the first part of the talk, we will present a minimax dynamic regret optimal policy under two somewhat strong assumptions: (i) the noise process is independent across time steps, and (ii) the total variation of the dynamics over T time steps is sublinear in T (we do not assume this variation is known). In the second part, we will relax both these assumptions. Since dynamic regret minimization is too strong a goal, we propose a policy that guarantees bounded-input-bounded-output stability in the closed loop. The talk will highlight how different perspectives for studying online control under non-stationary dynamics lead to novel statistical and algorithmic questions.

3 - Asymptotically Efficient Distributed Experimentation

Ankur Mani, University of Minnesota - Twin Cities, Minneapolis, MN, United States

Sequential decision making by a large set of independent agents who are not interested in experimentation, either because they are short lived or because experimentation is costly, has gained significant attention over the past decade. Even a little amount of experimentation from a few agents would benefit all others but coordinating such experimentation is challenging for a central planner. Academic literature has focused on mechanisms for promoting experimentation through monetary incentives and persuasion through careful information disclosure. We study a simple control that the central planner can use to coordinate experimentation. We consider a set of agents that observe their own history but not the histories of other agents. The central planner however can observe the history of all agents. In a continuous-time stochastic multi-armed bandit model, myopic agents observe their history, pick an arm and receive a reward at each time instant. In the proposed class of policies, the central planner, using the information it knows, at times unknown apriori, removes arms irrevocably that it considers inferior to other arms with sufficient confidence, thus forcing the agents to choose from a smaller set of remaining arms. We show that an appropriately chosen policy within this class can generate the needed experimentation and match the regret bounds for a centralized problem thus mitigating the cost of decentralization. We also quantify the minimum number of agents that are needed for such a policy to be asymptotically optimal and the impact of the number of agents on the speed of learning.

SA20

Summit - 340

Game Theory and its Applications

Invited Session

Decision Analysis Society

Chair: Shima Mohebbi, George Mason University, Fairfax, VA, United States

Co-Chair: Behnam Momeni, George Mason University, Fairfax, VA, United States

1 - Defense Critical Supply Chain Networks and Risk Management with the Inclusion of Labor: Dynamics and Quantification of Performance and the Ranking of Nodes and Links

Anna Nagurney, University of Massachusetts Amherst, Amherst, MA, United States

The efficient and effective performance of defense critical supply chain networks is essential to both national and global security. Disruptions to supply chains, heightened in the COVID-10 pandemic, and now further exacerbated because of growing geopolitical and other risks, as well as Russia’s war against Ukraine, have garnered the attention of decision-makers and policymakers, including those in the defense sector. In the paper, a rigorous methodological framework is presented for defense critical supply chain networks in the form of a defense supply chain network economy that captures the behavior of defense firms, which care about revenues as well as risk, and which includes the important labor resources and associated constraints. Variational inequality theory is used to provide alternative formulations of the governing

Nash Equilibrium conditions, with a dynamic model counterpart used for the construction of an easy to implement algorithm that yields closed form expressions at each iteration of the defense product path flows and the Lagrange multipliers associated with the bounds on labor hours available on supply chain links. A defense supply chain network efficiency /performance measure is proposed and an associated importance indicator for supply chain network components. A resilience measure is also given that quantifies the resilience of the defense supply chain network economy to disruptions in labor. The modeling and algorithmic framework, as well as the measures proposed, are then illustrated via numerical examples.

2 - The Weak Core, Partition-Based Universal Stability, and Their Risk Associations Through a Partial Order

Jian Yang, Rutgers University, Newark, NJ, United States

We are concerned with the stability of a coalitional game, i.e., a transferable-utility cooperative game. First, the concept of core can be weakened so that the blocking of changes is limited to only those with multilateral backings. This principle of consensual blocking, as well as the traditional core-defining one of unilateral blocking and one straddling in between, can then be applied to partition-allocation pairs. Each such pair is made up of a partition of the grand coalition and a corresponding allocation vector whose components are efficient and individually rational for the various constituent coalitions of the given partition. Among the resulting stability concepts, two are universal in that any game, no matter how ``poor'' it is, has its fair share of stable solutions. For a game possessing strictly positive values, furthermore, its imputations possess fractional interpretations. These would allow a certain ranking between games, which we deem as in the sense of ``centripetality'', to imply a clearly describable shift in the games' stable solutions. When coalitions' values are built on both random outcomes and a common positively homogeneous reward function characterizing players' enjoyments from their shares, this comparative static could help explain why it so often happens that aversion to risk promotes cooperation.

3 - Scalar Equilibria for Normal Form Games

Herbert Corley, The University of Texas at Arlington, Arlington, TX, United States

A scalar equilibrium (SE) is an alternative type of equilibrium usually in pure strategies for an n-person normal-form game G. It is defined using optimization techniques to obtain a pure strategy for each player of G by maximizing an appropriate utility function over the acceptable joint actions. The players' actions are determined by the choice of the utility function. Such a utility function could be agreed upon by the players or chosen by an arbitrator. An SE is an equilibrium in the sense that no players of G can increase the value of this utility function by changing their strategies. SEs are formally defined, and examples are given. In a greedy SE, the goal is to assign actions to the players giving them the largest individual payoffs jointly possible. In a weighted SE, each player is assigned weights modeling the degree to which he helps every player, including himself, achieve as large a payoff as jointly possible. In a compromise SE, each player wants a fair payoff for a reasonable interpretation of fairness. In a parity SE, the players want their payoffs to be as nearly equal as jointly possible. Finally, a satisficing SE achieves a personal target payoff value for each player. The vector payoffs associated with each of these SEs are shown to be Pareto optimal among all such acceptable vectors, as well as computationally tractable.

4 - Stochastic Differential Games for Designing Water Conservation Incentives

Behnam Momeni, George Mason University, Fairfax, VA, United States, Shima Mohebbi

Global concerns about water scarcity are intensifying due to poor freshwater management. This study develops stochastic differential games to obtain dynamic incentive strategies for water users located throughout river networks. Two networks of water users are considered: those who utilize groundwater and those who rely on surface water for agricultural purposes. A non-governmental organization (NGO) participates as another key player, providing conservation incentives to encourage water users to reduce their consumption. The NGO aims to offer an incentive scheme to each water user based on their unique characteristics, while addressing potential scalability challenges that may arise with an increasing number of players. Stochastic elements are captured by incorporating a Standard Wiener process into the state variables (volume of water in river and Aquifer). The Hamilton-Jacobi-Bellman equation and Ito lemma are then employed to determine the optimal decision variables for the water users and to analyze how the mean and variance of the state variables evolve when following the optimal decisions. Finally, an algorithm is proposed to solve large-scale convex and non-convex optimization problems. The Red River, the second-largest basin in the south-central United States, serves as the case study for the proposed model. Keywords: Stochastic decision games, Incentives, Freshwater management

SA21

Summit - 341

Emerging Technologies in Operations Management

Invited Session

Decision Analysis Society

Chair: Amir Hossein Moadab, Washington State University, Pullman, WA, 99163, United States

1 - Optimizing Drone Depot Location and Routing for Last-Mile Delivery: A Novel Application and Extension of the Black and White Traveling Salesman Problem

Juan Zhang, University of Wisconsin - Eau Claire, Eau Claire, WI, United States, Ozgur Kabadurmus, Haitao Li

As drone technology continues to revolutionize last-mile delivery, the challenges of limited flight range and payload capacity necessitate innovative solutions for efficient logistics. This research aims to address the critical question of how to strategically locate drone depots to mitigate the impact of constrained drone capabilities and optimally schedule drone routes to minimize drone facility and delivery costs. A location-routing optimization model is developed, which leverages and extends the unique modeling of the Black and White Traveling Salesman Problem. A case study for last mile healthcare delivery illustrates the usefulness of the model. This optimization framework has the potential to enhance the overall performance and scalability of drone delivery systems, making them more viable and effective in real-world applications.

2 - Unveiling the Influence of Blockchain Technology on ESG Indexes Across Supply Chains: An Empirical Investigation

Sahar Bajgani, WORCESTER POLYTECHNIC INSTITUTE, Worcester, MA, United States, Sara Saberi

This research project critically examines the effects of Blockchain Technology (BCT) on Environmental, Social, and Governance (ESG) metrics and financial indices within firms that have adopted this technology. The aim is to establish whether the integration of BCT leads to improved sustainability practices and enhances financial performance.

Quantitative data from a range of industries will be analyzed to compare the pre- and post-adoption performance of companies that have implemented blockchain solutions. This analysis will help isolate BCT's impact from other variables. Additionally, qualitative assessments will explore how the intrinsic features of blockchain—such as increased transparency, enhanced security, and decentralization—contribute to variations in ESG scores and financial outcomes.

The project will also investigate the potential of blockchain to facilitate compliance with regulatory standards and to foster trust among stakeholders, which are critical components of ESG criteria. Preliminary findings suggest that BCT can significantly influence corporate governance practices and operational efficiencies.

3 - Dispatching Drones after a Disaster Using a Team Orienteering Approach

Amir Hossein Moadab, Washington State University, Pullman, WA, United States, Chuck Munson, Arman Hosseini

We model the use of drones to aid in locating victims following a major disaster. Employing a two-phase approach, we integrate k-means clustering to identify potential hub locations. In the first phase, we optimize hub selection and drone allocation to each hub using a heuristic algorithm. The second phase focuses on efficient routing to ensure timely coverage, thereby enhancing search and rescue operations. We tested this model in three scenarios considering no threshold coverage for clusters, a specific percentage of threshold, and maximal covering hub selection. We identify important trade-offs for decision makers to consider when planning for and responding to major disasters.

4 - Reducing Search Advertisement Spending Through a Two-Stage Term Embedding and Optimization Approach

Vasileios Pavlopoulos, The University of Alabama in Huntsville, Huntsville, AL, United States, Hieu Pham, Ravi Patnayakuni

In the realm of e-commerce, efficient allocation of advertising budgets is crucial for maximizing profitability. This paper presents a novel approach to reducing search advertisement spend by leveraging negative exact matching of search terms that do not lead to conversions. Traditional methods often rely on waiting for a predetermined number of searches without clicks before negative-exacting a search term, which can be time-consuming and inefficient. To address this issue, we introduce a similar text-matching algorithm that utilizes word embeddings to identify similar queries, enabling proactive negative-exacting before incurring significant ad costs.

SA22

Summit - 342

Agricultural and Food Supply Chain Decision-Making for Societal Good

Invited Session

Decision Analysis Society

Chair: Deniz Berfin Karakoc, University of Illinois at Urbana-Champaign, Urbana, IL, United States

1 - A Stochastic Dynamic Approach for Task Assignment and Volunteer-Based Delivery Routing in Non-Profit Food Rescue Operations

Leila Hajibabai, North Carolina State University, Raleigh, NC, United States, Mehr Salami, Kuangying Li, Ali Hajbabaie

This research develops a stochastic dynamic program to optimize food rescue logistics for non-profit organizations. It aims to reduce hunger and food waste by maximizing demand coverage and minimizing routing costs, thereby improving the efficiency of food distribution networks. Characterized by fluctuating demand, variable volunteer availability, and constrained transportation resources, the problem suffers from extensive action space and highly random requests due to stochastic dynamics inherent in food rescue operations. A Markov decision process is employed that is enhanced by a Monte Carlo tree search technique to effectively navigate the stochasticities. The proposed integrated methodology dynamically assigns delivery tasks and designs optimal routes for delivery volunteers, incorporating new requests into the solution in real-time. To tackle the complexities due to the extensive action space, a set of heuristics is implemented that help prioritize and sample promising actions, enhancing the operational effectiveness of the proposed tree search algorithm. The effectiveness of the proposed approach is evaluated through comparative experiments using real-world data obtained from a collaborating food rescue agency. The preliminary results indicate that our methodology significantly outperforms existing state-of-the-art solutions in terms of operational efficiency and robustness. It not only improves the distribution of perishable food items to agencies and food-insecure households but also offers a scalable solution that can be adapted to other domains facing similar logistical challenges.

2 - The Many Faces of Equity? Food Bank Operations Through a Multi-Dimensional Equity Lens

Naurin Zoha, University of Michigan, Ann Arbor, MI, United States, Tanzid Hasnain, Julie Ivy

According to Feeding America, there are persistent disparities in food insecurity levels among different racial and ethnic groups in the US. In 2021, while about 10.2% of the US population were food insecure, 19.5% of the African American and 16.2% of the Hispanic populations were food insecure, compared to 7% of the White population. The level of food insecurity also varies among different population groups with different socio-demographic dimensions. As food banks develop distribution policies, they should be aware of the disparities in food insecurity that may exist among different demographics and may adversely impact equitable distribution. This study develops a single-period deterministic network flow model to explore the impact of different equity measures across multiple dimensions, namely geographic and demographic, on food bank's distribution in the presence of effectiveness (undistributed food minimization) criterion under efficiency (distribution cost minimization) constraint. We apply the model to our partner food bank's data to (i) explore the trade-off between dimensions of equity, and (ii) identify policy insights.

3 - Addressing Barriers to Food Access: A Comprehensive Study of Fresh Mobile Market Distribution in Low-Income Communities

Motunrayo Ogunmola, North Carolina A&T State University, Greensboro, NC, United States, Shona Morgan, Dwight Lewis, Lauren Davis

Food insecurity remains a global public health obstacle, highlighting inequalities vulnerable populations face. Using data collected from participants at several fresh mobile market (FMM) locations in a North Carolina (USA) county, combined with other US Census and geospatial datasets, we performed a case study examining FMMs' impact on their beneficiaries and serviced neighborhoods. More specifically, we examined the commute patterns of participants who visited FMMs and the sociodemographic characteristics of participants' residences and mobile market-serviced communities. Lastly, using contemporary spatial modeling, we enumerate the improvements in physical access to food assistance programs by looking at the impact of current FMM scheduling given the spatial distribution of existing food banks and pantries. To address remaining disparities in access to food assistance, we propose a spatial optimization approach for strategically positioning FMMs to minimize the travel burden to food assistance among vulnerable populations. The findings in this research can guide the strategic placement of FMMs in low-income communities tailored to their target communities' specific needs and constraints.

4 - Mapping Agri-Food Distribution onto Real-World Transportation Infrastructure in the United States

Deniz Berfin Karakoc, Arizona State University, Tempe, AZ, United States

Agricultural and food supply chains within the United States are essential for both global and local food security. These complex systems are subject to threats imposed by various economic, political, environmental, and infrastructural factors. Here, we focus on the transportation of agri-food commodities, which is a less well-studied feature, but necessary for connecting production to consumption. In this study, we develop a unified framework of data analytics, network science, and geographic information science techniques to map agri-food distribution within the United States onto real-world highway, railway, and waterway networks. We evaluate how the agri-food load on each transportation mode varies by the distribution of domestic, import, and export flows within the U.S. We also quantify the trade-off between economic efficiency, adaptability, and sustainability of agri-food distribution across transportation modes. We conclude that highways enable the greatest re-routing capacity for agri-food distribution, but it comes at the highest cost and carbon emissions; whereas waterways have the lowest cost and carbon emissions, but with the lowest adaptive capacity to disturbance. Our findings can inform efforts to balance affordability, resilience, and sustainability in agricultural and food transportation.

SA23

Summit - 343

Multiobjective Optimization Under Uncertainty

Invited Session

Multi Criteria Decision Making

Chair: Margaret Wiecek, Clemson University, Clemson, SC, United States

Co-Chair: Akshita Gupta, Purdue University, West Lafayette, IN, United States

1 - Two-Stage Stochastic Multiobjective Linear Programs and Their Properties

Akshita Gupta, Purdue University, West Lafayette, IN, United States, Susan Hunter, Margaret Wiecek

To address decision making under conflict and uncertainty, we consider a two-stage stochastic multiobjective linear program (TSSMOLP). As a generalization of its single-objective counterpart, the two-stage stochastic linear program, the TSSMOLP may have more than one conflicting objective in each stage. Thus, the second stage is modeled with a set-valued map resulting in a nondominated set rather than a real optimal value that is provided in a single-objective setting. In the first stage, the global Pareto set for the TSSMOLP is the set of nondominated points belonging to the union of the translated expected second-stage nondominated sets. Solving the TSSMOLP requires (a) proving properties of the TSSMOLP and (b) developing a theory to estimate the expected value of the second-stage nondominated set. We discuss structural properties of TSSMOLPs and recent theoretical advances for solving TSSMOLPs using a sample average approximation framework.

2 - Confidence Regions for Multi-Objective Stochastic Convex Programs with Uncertain Parameters

Susan Hunter, Purdue University, West Lafayette, IN, United States, Ziyu Liu, Raghu Pasupathy

We consider the context of constructing confidence regions on the efficient and Pareto sets of multi-objective stochastic convex programs with uncertain parameters. We first derive an asymptotically exact central limit theorem in the parameter space. Then, using the multiple objective functions as push-forwards, we construct asymptotically valid confidence regions on the efficient and Pareto sets. We illustrate projections of the confidence regions in the decision and objective spaces through an example.

3 - Upper Bounds on Scalar and Set-Based Regret for Online Multiobjective Optimization

Kristen Savary, Clemson University, Davenport, FL, United States, Margaret Wiecek

In online multiobjective optimization (OMO), solutions are computed before the objective functions are revealed. As a result, these solutions are not Pareto for the associated offline multiobjective problem and incur a cost, measured by regret.

Scalar regret and set-based regret are defined for OMO and used to evaluate the performance of the Online Gradient Descent Algorithm (OGDA). Using the scalarized OMO problem, the OGDA computes a single solution in every iteration and its success is measured by the scalar regret. Minimizing the generational distance associated with the OMO problem, the OGDA computes a set of solutions at every iteration and its success is measured by the set-based regret. A theoretical and numerical comparison of the upper bounds on the two regrets is presented.

SA24

Summit - 344

Optimizing Environmental Management: Simulation & Bioeconomics

Contributed Session

Contributed

Chair: Aoran Cheng, The University of Hong Kong, Rm17A02, The Met Azure, No.8 Liu To Road, Tsing Yi, Kwai Tsing, New Terrorities, HKSAR, Hong Kong, 999077, Hong Kong

1 - *Wildfire Rescue Missions Utilize Air Drone Carriers Planning*

Aoran Cheng, The University of Hong Kong, Hong Kong, Hong Kong, Shijie Pan, Yiqi Sun, Yulun Zhou, Kai Kang, Cristobal Pais, Zuo-Jun Shen

Preventing wildfires is an extremely important task and safeguard measure in this era. The primary objective of current forest fire management is to find effective methods to extinguish wildfires. However, when facing severe wildfires, we still lack an effective approach that can efficiently suppress the wildfire while ensuring the safety of forest firefighters. Therefore, this research explores the deployment, allocation, and near-optimal mechanism design of emerging drone swarms, as a substitution of traditional human based-firefighting, in the early stages of wildfire outbreaks. Specifically, we aim to optimize the path planning of each individual base during their firefighting missions to achieve the shortest possible time for wildfire suppression. Next, we will utilize geographic information projection and optimization methods to effectively allocate resources and devise scheduling strategies to address the coordination and scheduling challenges of limited resources and time-constrained drones and trucks. Through extensive research and practical experience, our goal is to enhance the efficiency promise of task execution. Furthermore, by integrating optimal strategies for individual drones and drawing inspiration from collective behaviors observed in nature, we could improve the overall firefighting efficiency of the drone swarm. Based on these models, we will explore the practical application of the proposed approach in real firefighting scenarios, providing innovative solutions for wildfire prevention and control.

2 - Enhanced Modeling of Fan Speed Effects on Greenhouse Climate and Agricultural Product Drying

Linda Orjaroen, Chiang Mai University, Chiang Mai, Thailand, Chulin Likasiri

This study advances latest research on the effectiveness of solar greenhouse drying, concentrating on a refined dynamic model that simulates the impact of fan speed adjustments on internal greenhouse conditions—temperature, humidity—and the drying kinetics of bananas. Utilizing comprehensive data from an ongoing banana drying case study, including measurements of air temperature, humidity, irradiance, and the temperature and weight of the bananas, the model employs advanced Ordinary Differential Equations (ODEs). The equations are coupled with sophisticated unconstrained optimization methods to enhance the empirical fit. Enhanced correlation analysis explores the intricate relationships between fan speed adjustments and key environmental variables. All parameters have been rigorously validated using root mean squared error minimization techniques to ensure the model's accuracy and reliability. The refined mathematical model provides more precise predictions of the final weight of dried bananas and recommends optimal fan speed settings to improve drying efficiency. This focused project contributes significantly to optimizing greenhouse drying processes for bananas and promoting more efficient and sustainable local agricultural practices.

3 - A framework to identify and prioritize key performance indicators: assessment of the effectiveness of inclusive climate actions for nature-based solutions

Sara Torabi, Politecnico di Torino, Turin, Italy

Nature-based Solutions (NbS) are increasingly acknowledged as vital strategies against climate change. However, citizens and ecosystems benefit differently from the functions that NbS provide. To better account for NbS trade-offs, theory and practice have embraced Inclusive Climate Actions (ICAs) that simultaneously tackle climate change and urban inequalities. Currently, there is a notable gap in a comprehensive assessment of the environmental, economic, and social implications of ICAs for NbS. This research develops a novel evaluation and monitoring (E&M) framework to identify Key Performance Indicators (KPIs) for evaluating the effectiveness of ICAs for NbS. The methodology is structured into three distinct phases: (i) identification, (ii) refinement, and (iii) prioritization. This approach integrates both quantitative and qualitative techniques and actively involves stakeholders in the process. A set of 13 KPIs is proposed consisting of 2 economic, 5 environmental, and 6 social indicators was identified. Furthermore, these results can be further enhanced engaging local stakeholders to select, when necessary, supporting indicators that are tailored to address specific contextual challenges. The E&M framework assesses the effectiveness of completed or ongoing projects, or to design and evaluate future scenarios. Results will help cities to implement ICAs that incorporate fairness and distribute NbS impacts as equitably as possible. This step is pivotal for empirically evaluating the effectiveness, design, and planning of future scenarios of ICAs within the Urban Living Labs of five European cities. This study is part of the broader European project GREEN-INC, growing effective and equitable nature-based solutions through inclusive climate actions.

4 - Pareto Optimization for Selecting Discriminating Test Locations in Plant Breeding

Mohammadreza Kiaghadi, Iowa State University, Ames, IA, United States, Sigurdur Olafsson

Developing new and better cultivars of food crops is an important challenge that requires expensive and time-consuming field trials. In early-stage experiments thousands of experimental cultivars may be planted in a small number of locations with the goal of only continuing trials with only a small fraction of the very best cultivars in the next season. The selection of experimental planting locations thus plays an important role in plant breeding, and in early-stage experiments a key consideration is to include locations that have the best ability to discriminate the main genotype effects of cultivar, that is, to screen the potential high performers from the average to poor cultivars. Traditionally, this discriminative ability of locations has focused on precision, but recent work also suggests that some locations may be more sensitive to changes in the main genotype effect, that is, have what we will refer to as higher relative discriminative value. We show that while these are both valuable, they are competing measures of discriminative ability, that is, that no location will maximize both precision and relative discriminative value, and there is therefore a tradeoff that should be considered by plant breeders. We further show how this tradeoff can be addressed by proposing how a set of discriminating location can be constructed using ideas based on pareto optimization.

5 - When does green technology diversification affect firms in catching up with industry leaders? The role of environmental policy coordination

yunze li, Harbin Engineering University, Harbin, China, People's Republic of, Weiwei Liu, Ruby Lee

Inspired by the conference calling for more research on smarter decisions for a better world and organizational learning theory, we examine green technology (GT) diversification among firms from an emerging economy (EE) and propose that an increase in GT diversification can facilitate EE firms to catch up with a global industry leader in different ways. GT diversification reflects the number of GT domains out of seven scopes classified by the World Intellectual Property Organization, and thus, the more GT domains an EE firm has, the more diverse it is. We theorize that when GT diversification increases, learning from a wide range of knowledge enables EE firms to catch up with the industry leader's economic performance more closely. However, owing to a wider range of environmental regulations connected to various scopes of GT, compliance with many regulations can slow down EE firms' learning and catch-up with industry leaders. We further explore the role of environmental policy coordination. Using textual analysis of public policies, we develop a novel measure to capture the extent of environmental policy coordination between government departments. We further collect longitudinal data (2012-2023) from publicly traded firms in China to examine the effect of GT diversification and find that GT diversification affects economic catch-up performance positively but environmental catch-up performance negatively. However, when a province's environmental policy between government units is more coordinated, it attenuates the effects of GT diversification on catch-up performance in different ways. Our findings provide novel implications for theory and practice.

6 - Revisiting the Impact of Renewable Energy on PPAs

Subas Acharya, Johns Hopkins University, Baltimore, MD, United States, Benoit Chevalier-Roignant, Helyette Geman

The rise of green businesses and renewable energy marks a pivotal shift in the global pursuit of net zero emissions. Power Purchase Agreements (PPAs) allow corporations and businesses to source green energy. In the long run, the larger share of cost-effective renewable sources in the overall power generation mix will lead to lower electricity prices. This price dynamics will prompt offtakers to withdraw from a PPA signed at a time where the energy producer could charge a premium on green energy and the generation of solar and wind was limited. Assuming a mean reverting price process with a decreasing level of mean reversion, we analyze the date at which an offtaker is likely to renegotiate on a PPA and key statistics on this stopping time to understand better the PPA average duration and revenues producers can expect when defining the contract price at date 0.

SA25

Summit - 345

Creative Applications of AI

Contributed Session

Contributed

Chair: Devansh Jalota, Stanford University, Stanford, CA, United States

1 - Exploring the Role of Generative AI in Transforming the Analytics Education in Business School

Qizhang Liu, National University of Singapore, Singapore, Singapore

The emergence of generative AI technologies, exemplified by innovative platforms like ChatGPT, has catalyzed widespread disruption across various sectors, with education standing as no exception. Educators worldwide are actively engaged in discussions, brainstorming sessions, and experimental endeavors, seeking short-term mitigative measures and long-term adaptation strategies in response to the transformative impact of generative AIs. However, despite the palpable urgency surrounding this issue, the existing scholarly discourse on the efficacy of such measures or strategies remains notably sparse.

This study presents a comprehensive analysis of the transformative interventions implemented within an analytics course at the Business School of the National University of Singapore. Spanning two academic semesters, these interventions encompassed a spectrum of modifications, including the redefinition of teaching objectives, a paradigm shift in teaching philosophy, and the overhaul of assessment methodologies. The efficacy of these interventions was assessed through surveys administered to students, revealing a discernibly positive reception to the implemented changes.

In light of these findings, a series of proactive recommendations are proffered to circumvent the potential pitfalls associated with overreliance on generative AI technologies. These recommendations serve as pragmatic guidelines for educators seeking to harness the transformative potential of generative AIs while concurrently fostering the cultivation of essential analytical competencies among students.

2 - Talk Me Into It: Evaluating Advertising Through GenAI Chatbots

Parto Tavassoli, Simon Fraser University, Burnaby, BC, Canada, AIshwarya Shukla, JM Goh

The introduction of Generative AI technologies has significantly transformed digital interactions, with profound implications for marketing practices. Every day, millions of users across the globe engage with GenAI chatbots, highlighting their widespread popularity and significant potential. Recently, both marketers and AI development firms have recognized the potential of GenAI chatbots as a novel medium for reaching potential customers through personalized and interactive communication. In this study, we aim to examine the implications of embedding advertisements within interactions with GenAI chatbots. We propose that such advertisements may diminish trust in the chatbots' responses, as they may be perceived to be biased toward sponsored products. Additionally, the perception of brands utilizing this marketing strategy could be adversely affected, especially if the presence of advertisements is not transparently disclosed to users. These factors may influence user engagement metrics, such as *click-through rates*, which entails further investigation into how people react to this advertising approach. We will also explore the moderating effect of *task complexity* on these potential negative outcomes. Our hypotheses suggest that higher search costs and perceived complexity in finding the right products might mitigate users' sensitivity to receiving sponsored recommendations from these chatbots. We propose that by nudging users about the complexity of searching, people may become more likely to be receptive to the chatbot's suggestions when they include sponsored material. The proposed hypotheses will be empirically tested through a lab experiment in subsequent phases of this research.

3 - Understanding Human-AI Collaboration in Generative AI Text Detection

Fairy Gandhi, University of Colorado Boulder, Boulder, CO, United States, David Dobolyi, Julian Lehmann, Matthew McCarthy

Recent advancements in generative artificial intelligence have led to a surge in machine-synthesized content in a wide array of fields. AI-generated text has often proven indistinguishable from human-written text, presenting challenges across industries. Although much focus has been on distinguishing text as either human- or AI-generated, many instances involve a blend of both human and AI contributions. AI-text-detection tools, while showing promise in identifying AI-generated content, struggle to detect mixtures of human and AI writing. Humans collaborating with such tools may misplace their trust in the tools' recommendations during their decision-making process. We aim to understand the influence of AI-text-detection tools on human decision-making in a three-way text classification task. First, through a large-scale, classroom-based field study, we collected prompt-based student responses that were self-reported as human-written, AI-written, or a mixture of both. Second, we are currently conducting an online experiment that has participants evaluating these student responses with or without the assistance of a real-world, commercial AI-text-detection tool. We hypothesize that individuals collaborating with recommendations from the tool will generally outperform those without it. Critically, however, our findings from the field study suggest that accuracy will be higher only in the human-written and AI-written conditions when relying on the tool's recommendations, whereas accuracy will be significantly lower in the mixed condition due to flawed tool recommendations present in the mixed condition. This research will provide important insights into the augmented human/AI decision-making process in education, which will apply to other settings such as content moderation, cybersecurity, and beyond.

4 - When Simple is Near-Optimal in Security Games

Devansh Jalota, Stanford University, Stanford, CA, United States, Michael Ostrovsky, Marco Pavone

Fraudulent or illegal activities are ubiquitous across many applications and involve users bypassing the rule of law, often with the strategic aim of obtaining some benefit that would otherwise be unattainable within the bounds of lawful conduct. However, user fraud is detrimental, as it may compromise safety or impose disproportionate negative externalities on particular population groups.

To mitigate the potential harms of users engaging in fraudulent activities, we study the problem of policing such fraud as a security game between an administrator and users. For this security game, we study both welfare and revenue maximization administrator objectives. In both settings, we show that computing the optimal administrator strategy is NP-hard and develop natural greedy algorithm variants for the respective settings with constant factor approximation ratio and resource augmentation guarantees.

We further study several model extensions, including incorporating contracts into our framework and generalizing our model to incorporate additional constraints beyond an overall resource constraint. While greedy approaches do not, in general, work well for arbitrary constraints, we identify sufficient conditions on the constraint structure under which our proposed greedy algorithms can still achieve constant factor approximation ratios to the optimal policy.

Finally, we present experiments based on real-world data obtained from Stanford University's Department of Public Health and Safety on monitoring parking violations on Stanford's campus. Our results demonstrate that our proposed algorithms can increase campus revenues from parking permit purchases by over $300,000 every year relative to the status-quo parking enforcement policy.

SA26

Summit - 346

Managing in High Technology Environments

Invited Session

New Product Development

Chair: John Angelis, University of Lynchburg, Lynchburg, VA, United States

Co-Chair: Juliana Hsuan, Copenhagen Business School, Frederiksberg, DK-2000, Denmark

1 - The Effects of Blame Shifting and Shopping Intensity on Revisit Likelihood in Data Breaches

John Angelis, University of Lynchburg, Lynchburg, VA, United States, Rajendran Murthy

Data thefts have increased in frequency and cause a loss of customer trust and business. Previous research has studied how the public assigns blame for the crime, and how different individuality levels of data breaches affect revisit intention. In addition, the negative emotional response of the public to data breaches has been frequently studied, finding that fear is highly correlated with revisit intentions. However, previous research has been inconclusive on whether frequent shoppers will be more likely to shift blame away from the business, and whether they will be more likely to return. Our survey of 315 individuals provides insight into how frequent customers differ from the general public. Specifically, the relationship between placing greater proportion of blame on data breach victims (relative to the hacked business) and revisit intention is moderated by prior shopping intensity. Those who placed greater blame on the data breach victims were more likely to return to the breached business. Frequent shoppers are indeed more likely to return to the breached business, indicating that their previous experience does affect their revisit intention. We put this result in context with the literature on customer loyalty, and recommend that businesses monitor public sentiment carefully and match their crisis response to their target customer audience.

2 - On Syntactic and Semantic Attention Summary (SSAS): An Approach to Improve Summary Generation through LLM's

Nitin Mayande, Tellagence, Naperville, IL, United States, Sharookh Daruwalla, Sumedh Khodke, Nitin Joglekar, Charles Weber

In today's information age, the ability to efficiently extract key points from vast amounts of text is crucial. Automatic summarization, a subfield of NLP, tackles this challenge by generating condensed versions of documents while preserving essential information. However, achieving accurate summaries requires models to go beyond simple word frequencies.

Accurate summarization hinges on effectively bridging the gap between syntax and semantics. Syntactic cues often point towards semantically relevant information. Syntactic alignment provides a roadmap, identifying where to look for important information within a sentence structure. Semantic alignment then interprets the identified elements, extracting their meaning and relationships. In this paper we propose a combined approach leads to a more nuanced understanding of the text, enabling the generation of summaries that are not just grammatically sound but also semantically accurate. It also ensures that the broader context of the dataset is captured.

3 - Real-Time Product Development using Generative Artificial Intelligence

Charles Weber, Portland State University, Portland, OR, United States, Antonie Jetter, Dahm Mongkol Hongchai, Ameeta Agrawal, Yufei Tao

A workshop was conducted to train practitioners in real-time new product development (NPD) utilizing AI. Three teams of managers and engineers were allotted six hours to develop a marketing plan prototype using ChatGPT and other AI tools. The plans included customer identification, market analysis, persona creation, and a 15-minute presentation aimed at inspiring a sales force. A panel of judges, including the authors and an external marketing professional, assessed the teams' strengths, weaknesses, and overall viability. The workshop validated the efficacy of AI in NPD across various industries, with the outputs deemed "highly creative" and potentially viable. The primary takeaway was the demonstrated acceleration of product development activities.

4 - Five Pivotal Practices for Leaders to Cultivate Empathy within Virtual Scrum Teams

Dahm Mongkol Hongchai, Portland State University, Portland, OR, United States, Charles Weber

In today's digital and global economy, it's essential for business leaders to possess not only business and technology skills but also empathy. Yet, there is a significant lack of research focusing on the role of empathy within Agile product development, particularly when using the Scrum Framework. A qualitative, empirical study conducted at a professional accounting firm uncovered critical insights into how leaders can foster empathy within virtual Scrum teams. Actionable approaches to leadership involve five themes: Scrum fundamentals, meditation, self- and other-analysis, healthy interactions, and informal connections.

5 - Measuring servitization configurations through modularity

Juliana Hsuan, Copenhagen Business School, Frederiksberg, Denmark

Servitization is enabled through product and service configurations to offer value propositions that are attractive to customers. Modularity can provide powerful insights on how to measure the configuration options and respective implications for business model innovations.

SA27

Summit - 347

Product and Business Model Innovation

Invited Session

Technology, Innovation Management and Entrepreneurship

Co-Chair: Niyazi Taneri, Cambridge Judge Business School, Cambridge, United Kingdom

1 - Leveraging Suppliers for Innovation: Sourcing and Investment Strategies

Pascale Crama, Singapore Management University, Singapore, Singapore, Gaoyan Lyu, Yi Xu

Manufacturers benefit from supplier innovation that creates value for their customers. Yet because suppliers only appropriate a fraction of the value created, their investment level may not match the manufacturer's preferences. We show that the manufacturer prefers single sourcing for highly profitable innovations but exploits the competition effect of dual sourcing for less profitable innovations. Furthermore, the manufacturer may opt to co-invest in supplier innovation to increase supplier effort. Interestingly, we find that even in the face of symmetric suppliers and under dual sourcing, the manufacturer may prefer to invest asymmetrically in the two suppliers. Finally, we find that the structure of the optimal sourcing and investment strategy differ depending on whether the innovation leads to product introduction or product improvement. Innovations that lead to new product introduction never lead to dual sourcing with asymmetric investment, whereas product improvement innovations not only favor dual sourcing with or without manufacturer investment, but may also lead to asymmetric investment under dual sourcing with symmetric suppliers.

2 - Unraveling the Implications of Silent Labor TIME (Slt) in the Gig Economy

Srishti Arora, INSEAD, Singapore, Singapore, Vivek Choudhary, Sameer Hasija

The gig economy has become integral to the global economy, driven by the labor flexibility it affords companies and the self-scheduling options available to workers. However, it also introduces an additional discretion for workers; they need to search for tasks during downtime, i.e., intervals when the platform does not assign them tasks. This period of uncompensated task-seeking, which we call “Silent Labor Time (SLT),” necessitates balancing effort between searching and executing tasks, impacting execution time. Our research aims to establish how the effort allocated during SLT affects workers’ performance and earnings and identify factors that moderate this relationship. In food delivery, the distance drivers travel to find the next order, called “relocation distance,” represents the effort allocated during SLT.

Collaborating with a food delivery platform, we find that, on average, drivers relocate 2.6 km before each order, and a km increase in relocation distance reduces order allocation by 5.4%, order speed by 2.7%, and earnings by 14.8% in the subsequent hour. The primary reason for this decline is drivers allocating significant effort to searching for tasks during SLT, subsequently conserving energy when executing tasks. Relocations that are not towards familiar clusters and reduce supply-demand balance are most detrimental to workers' performance and earnings. Our findings suggest that relocation adversely affects drivers' earnings and operational performance in subsequent orders, ultimately impacting the platform's efficiency. We offer actionable insights by suggesting strategies for the management of effort allocation during SLT by different driver groups.

3 - How Do Robots Affect Firms&Rsquo; Innovation Performance? Evidence from Spanish Manufacturers

Yiyao Zhou, UCL School of Management, London, United Kingdom, Bilal Gokpinar

This paper examines the impact brought by robot use on manufacturing firms' innovation performance. The analysis uses a rich panel dataset of Spanish manufacturing firms over 27 years (1990-2016). Our findings document, the first time in the literature, that robot use has a negative effect on firms' process innovation. However, we do not observe a similar effect on firms' product innovation. We also explore mechanisms by which robot use may affect process innovation. We find that the negative effect of robot use on process innovation is only salient for complex manufacturing, rather than light manufacturing or heavy manufacturing. In addition, we find that the negative effects brought by robots on process innovation are smaller for older firms. These results point to a potential mechanism whereby robots may impede process innovation through reducing human involvement. Our findings highlight possible disadvantages brought by robots in manufacturing firms, a notion neglected by the previous literature.

4 - Shifting Targets: How Funding Goal Changes Influence Donation-Based Crowdfunding Success

Xiahua Wei, University of Washington, Bothell, Bothell, WA, United States, Weijia You, Xiaoya Xie

In the dynamic world of crowdfunding, the ability to pivot and adjust fundraising goals can make or break a campaign. With limited research on this topic in the literature, our study delves into the intriguing phenomenon of modifying fundraising targets mid-campaign and its impact on donor behaviors. Utilizing panel data from China's largest donation-based crowdfunding platform, we uncover that announcing changes in funding goals significantly boosts donations on the same day, increasing both the number of participants and the average donation per person. Interestingly, increasing fundraising targets has a more pronounced effect on donor behaviors compared to reducing them. Text analysis further reveals distinct motivations behind these adjustments. Beyond immediate impacts, our study investigates the overall performance of projects with modified goals, finding they achieve higher completion rates, daily fundraising averages, and cumulative funds raised. These insights offer valuable guidance for fundraisers on setting and adjusting funding targets to enhance donation-based crowdfunding campaign success.

5 - The Effects of Robots on the Workplace

FNU Adrianto, University of Minnesota, Minneapolis, MN, United States, Avner Ben-Ner, Ainhoa Urtasun

This paper examines the effects of robots across various occupations in U.S. manufacturing plants, extending extant research conducted at the firm and industry levels. We use a difference-in-differences approach matched on industry, commuting zone, and plant size to estimate how employment and skill demand for various occupations change after robot adoption. We find that the introduction of robots is associated with a 150 percent increase in the number of job postings and an increase in employment of 15 percent; the increase is larger in production than in support jobs. Comparing effects across plants within adopting firms, we show that the expansion only occurs in the robotic plants, suggesting that prior firm-level studies overlooked the distinction between adopter and (majority) nonadopter plants within firms, underestimating the robotization effect. We find a negligible employment effect at the industry level as the positive effect in adopters is counterbalanced by the loss of workers in nonadopters. The majority of jobs do not change skill composition following the adoption, but the robotized part of the plant requires more design, production, maintenance, repair, and programming skills. We provide credible evidence that the productivity and robot-human complementarity effects dominate any displacement effect and that loss of employment is limited to outcompeted nonadopters.

SA28

Summit - 348

Recycling, Repurposing and Rebalancing in Precision Agriculture and Sustainable Systems

Invited Session

MSOM: Sustainable Operations

Chair: Metin Cakanyildirim, The University of Texas at Dallas, Richardson, TX, United States

Co-Chair: Lingling Shi, McMaster University, Hamilton, ON, Canada

1 - Circular Economy of EV Batteries: Economic and Environmental Impacts of Repurposing

Lingling Shi, McMaster University, Hamilton, ON, Canada, Metin Cakanyildirim, Sila Cetinkaya

The fast growth of electric vehicle (EV) and the increased energy storage markets stimulate the demand for batteries and in turn the critical minerals, which face high supply uncertainty. We investigate the economic and environmental impacts of repurposing spent EV batteries for energy storage in addition to recycling.

2 - Fighting the Plastic Pollution: Product Ban Regulation and Voluntary Compliance

Natalie Huang, University of Minnesota, Minneapolis, MN, United States, Wenli Xiao, Feifei Shan, Aditya Vedantam

We study a firm's voluntary recycling efforts as a proactive compliance strategy for future regulation to ban the sales of its products. We study the economic and environmental implications of the product ban regulation and the firm's compliance incentives.

3 - Equitable Supply-Demand Balancing: Tailoring Equitable Incentives for Demand Response Programs in Power Grids

Solmaz Abbaspour, University of Massachusetts, Amherst, MA, United States, Senay Solak

Demand response programs aim to shift electricity consumption from peak periods to off-peak periods to better match supply with demand, typically through prices or monetary incentives. We investigate equitable incentivization policies in demand response programs by considering diverse characteristics of various demographic groups. We model the trade-offs between equity and efficiency in such incentivization policies, and propose mechanisms that achieve efficiency while ensuring social equity among different customer profiles. The results are expected to provide utility companies with actionable insights for tailoring their demand response strategies more effectively, thus promoting a more sustainable and equitable energy management system.

4 - Evaluating the Trade-Offs Between Safety and Sustainability in a Supply Chain

Gal Raz, Ivey Business School at Western University, London, ON, Canada, Pantea Saremi, Jason Nguyen

This paper examines the impact of investing in safety for sustainable products. We consider a market with two competing manufacturers, one selling a conventional product, and the other a more environmentally sustainable one that could have some safety concerns as a result of altered material and/or production process. Examining the market, we evaluate the tradeoff between safety and sustainability for sustainable products and their economic and environmental impact. This evaluation entails examining sustainable product manufacturer's optimal strategy given the safety threshold they aim to attain.

5 - Outcome-based pricing for precision agriculture services

Heng Chen, University of Nebraska-Lincoln, Lincoln, NE, United States, Ying Zhang

Precision agriculture has been promoted by agricultural technology providers through the use of outcome-based pricing (OBP). In this paper, we examine the effects of OBP on the adoption rate of precision agriculture, and the benefits it offers to both farmers and service providers. We develop a two-period game model that incorporates providers' learning from experience. We also explore the implications of government intervention when the service provider switches to OBP from traditional pricing.

SA29

Summit - 420

Advances in Network Optimization & Design

Invited Session

OPT: Integer and Discrete Optimization

Chair: Luca Wrabetz, University of Pittsburgh, Pittsburgh, United States

1 - Exploiting players' sub-optimal strategies in multi-level interdiction games.

Rafal Muszalski, University at Buffalo, Buffalo, NY, United States, Jose Walteros, Prashant Sankaran

In a two-player interdiction game, an exploitative strategy consists of a set of moves conducted by the leader to exploit the follower's inability to play optimally. Exploitative strategies can arise when there are imbalances in the game setup across the players, such as when the follower, due to some technological limitation, is restricted to conducting heuristic moves instead of optimal ones. In this work, we formulate an exploitative framework for solving multi-level interdiction games, leveraging a hybrid deep learning optimization approach. Our framework implements a Graph Neural Network (GNN) component that predicts sub-game solutions based on the structural characteristics of the game, as well as historical observations of previous iterations of the game, including the players' previously used strategies. We show that, by sampling solutions generated by the GNN, our approach can generate an estimated characterization of the follower's decision space and use it to parameterize an optimization model tailored to produce exploitative strategies. We test the performance of our framework by solving several three-level interdiction games, including a shortest path interdiction problem with fortification.

2 - Simple Randomized Rounding for Max-Min Eigenvalue Augmentation

Haeseong Yang, University of Pittsburgh, Pittsburgh, PA, United States, Jourdain Lamperski, Oleg Prokopyev

We consider the max-min eigenvalue augmentation problem: given an n × n symmetric positive semidefinite matrix M , vectors a1, . . . , am ∈ R^n, and a positive integer k ≤ m, the goal is to choose a subset I ⊂ {1, . . . , m} of cardinality at most k to maximize the minimum eigenvalue of the matrix M + ∑_{I∈I} a_ia^⊤_i . The problem captures both the maximum algebraic connectivity augmentation and E-optimal design problems. There are two distinct research streams dedicated to these problems that develop approximation algorithms based on rounding optimal semidefinite programming relaxation solutions. Interestingly, both research streams empirically evaluate the performance of "simple" randomized rounding methods but do not establish theoretical approximation guarantees for them. In this paper we establish approximation guarantees for a simple randomized rounding method applied to the general max-min eigenvalue augmentation problem. Furthermore, we investigate its empirical performance through computational experiments.

3 - Commit or defer shortest path interdiction

Shuai Shao, Oklahoma State University, Stillwater, OK, United States, Juan Borrero, Oleg Prokopyev

We study commit or defer interdiction problems when the interdictor has partial initial information about the network and the evader has complete knowledge of the network, including its structure and arc costs. The interdictor has limited resources for interdiction which cannot be renewed after being used. In each time period, the interdictor chooses to block some arcs, and observes feedback from the shortest path used by the evader. By observing the evader's feedback, the interdictor improves her knowledge about the network and adjusts her strategy for the later time periods accordingly to maximize the cumulative cost incurred by the evader. We show that in the worst-case, for arbitrary sets of initial information, interdicting one arc at a time is an optimal worst-case policy. For specific sets of initial information, however, this is not true. For these cases we consider optimistic and pessimistic policies that balance performance and resource expenditures by using Pareto frontiers. We formulate and compute these policies using mixed-integer programming, and show that they can outperform other benchmark policies in our numerical experiments. In particular, these policies as many arcs as the resources permit whenever the interdictor has a reasonably good estimate of the costs of the network, and interdict only one arc whenever the estimates are not good.

4 - Adaptive shortest-path interdiction of multiple followers

Luca Wrabetz, University of Pittsburgh, Pittsburgh, PA, United States

In this work, we define and describe a novel interdiction problem, the Adaptive Shortest-Path Interdiction of Multiple Followers Problem. In this problem, a decision maker selects multiple policies (subsets of arcs to attack or remove) in order to robustly interdict multiple followers who have different costs on the network. We develop three exact algorithms, as well as an approximation algorithm for the problem.

SA30

Summit - 421

Modeling Problem Structure for Large Optimization Problems

Invited Session

OPT: Computational Optimization and Software

Chair: Alan King, IBM Research, Yorktown Heights, NY, United States

1 - Modeling Uncertainty in Mosdex

Alan King, IBM Research, Yorktown Heights, NY, United States

The Stochastic MPS format (S-MPS) is the go-to data format for Stochastic Programming, which applies optimization-based models and algorithms to address decision-making under uncertainty. This talk discusses the implementation of S-MPS in the MOSDEX framework.

2 - Representing Structured Optimization Models in Mosdex

Jeremy Bloom, self, Sunnyvale, CA, United States, Alan King, Matthew Saltzman, William Hart

MOSDEX (Mathematical Optimization Solver Data Exchange) is a new standard for representing optimization models that has many capabilities including independence from solvers, modeling languages, and programming languages, and end-to-end support of the application data flow. In particular, MOSDEX uses Structured Query Language (SQL) to specify compactly and efficiently the data transformations that feed the highly specialized artifacts that comprise an optimization model. Critically, existing formats poorly capture models' higher-level structure, limiting their usefulness in specialized domains such as stochastic programing and algorithms such as decomposition. This presentation presents extensions of MOSDEX to reflect such structure including: 1) modules to represent subproblems, 2) structure tables to represent relationships among modules, and 3) replication of modules on subsets of the data. The presentation will include explicit examples of their use in MOSDEX .

3 - Enabling Performant Optimization Modeling Languages

William Hart, Sandia National Laboratories, Berlin, MA, United States

We revisit the design of optimization modeling languages to reassess their role in the formulation and analysis of large, complex applications. Modeling languages like AMPL, Coek, GAMS, JuMP and Pyomo simplify the process of expressing complex mathematical expressions in an optimization formulation, and they automate the translation of these models to automate the execution of solvers and related model evaluations (e.g. gradient and Hessian calculations). Here, we consider the implications of modeling language design on the total computational cost for solving complex applications. Specifically, we discuss how a compact but restricted expression of complex equations can be exploited to enable nontrivial performance optimizations. Further, this idea supports the parallel generation of models, generation of LaTeX model descriptions, and interfaces to high-level representations like MOSDEX. We illustrate this idea using examples from Pyomo and Coek, and describe Smoek, a new python-based modeling language that explicitly exploits compact expressions.

4 - Mosdex for the Coin-or Open Solver Interface

Matthew Saltzman, Clemson University, Clemson, SC, United States

We describe a C++ library for reading and writing MOSDEX files from the COIN-OR Open Solver Interface (OSI). The library provides the capability of communicating problem instance data and structure to solvers supported by the OSI.

SA31

Summit - 422

Machine Learning for Optimization - I

Invited Session

OPT: Machine Learning

Chair: Gonzalo Constante Flores, Purdue University, West Lafayette, IN, United States

1 - Ensuring Neural Network Solution Feasibility for Constrained Optimization

Minghua Chen, City University of Hong Kong, Hong Kong, Hong Kong

There has been growing interest in employing neural network (NN) to directly solve constrained optimization problems with low run-time complexity. However, it is non-trivial to ensure NN solutions strictly satisfying problem constraints due to inherent NN prediction errors. Existing feasibility-ensuring methods either are computationally expensive or lack performance guarantee. In this talk, we present homeomorphic projection as a low-complexity scheme to guarantee NN solution feasibility for optimization over a general set homeomorphic to a unit ball, covering all compact convex sets and certain classes of nonconvex sets. The idea is to (i) learn a minimum distortion homeomorphic mapping between the constraint set and a unit ball using an invertible NN (INN), and then (ii) perform a simple bisection operation concerning the unit ball so that the INN-mapped final solution is feasible with respect to the constraint set with minor distortion-induced optimality loss. We prove the feasibility guarantee and bound the optimality loss under mild conditions. Simulation results, including those for non-convex AC-OPF problems in power grid operation, show that homeomorphic projection outperforms existing methods in solution feasibility and run-time complexity, while achieving similar optimality loss. We will also discuss open problems and future directions.

2 - Safe Learning Control of Inverter-Based Power Systems with Stability and Safety Constraints

Zixiao Ma, University of Washington, Seattle, WA, United States

As we strive for decarbonization, the power system is evolving to integrate more renewable energy resources, contributing to a cleaner and more sustainable energy landscape. Nevertheless, renewable resources such as wind and solar power typically exhibit low system inertia and have inherently variable output. Consequently, systems with a high penetration of renewable energy are more vulnerable to disturbances and fluctuations in power quality. To improve system reliability, it is crucial to control the critical states such as voltage, frequency, and current to stay within the safety constraints. Conventional computational methods require solving a constrained optimization problem to calculate the control signal iteratively. Such a method is computation-costly for high-order or nonlinear systems and is too slow to be applied in real-time. To overcome the above problems, I proposed a safety-critical voltage control method for inverter-dominated renewable energy systems to improve online computational efficiency while guaranteeing safety constraints. The key idea is to leverage the learning ability of neural networks to cast the online iterative constrained optimization into the offline training. Instead of using the neural network as a black box, we substitute it explicitly into the closed-loop system for deriving the stability and safety constraints. Then, the neural network is trained to optimize the control performance and the size of the stability region under the constraints. The well-trained controller is a static function of system states, such that it can be computed rapidly for online implementation.

3 - Ac-Network-Informed DC Optimal Power Flow for Electricity Markets

Gonzalo Constante Flores, Purdue University, West Lafayette, IN, United States, André Quisaguano, Antonio Conejo, Can Li

The DC optimal power flow (DC-OPF) problem is the workhorse for power systems and market operations in the power industry. The growth in electric demand due to data centers and electrification in the transportation and industrial sectors has driven the operation of power systems to conditions where the solution of the DC-OPF problem could be inaccurate with respect to its nonconvex counterpart. This talk presents a parametric quadratic approximation for the AC-OPF model that reduces the dispatch approximation error with respect to its nonconvex formulation while targeting desirable market properties. The parametric approximation preserves the physics-based formulation provided by the DC-OPF model and leverages the market and physics information encoded in the data-driven demand-dependent parameters. To enable the use of the considered model for real-time applications, we propose a supervised-learning approach to predict near-optimal parameters, given a certain metric with respect to the dispatch outcome and locational marginal prices. The training dataset is generated based on solutions of the AC-OPF problem and a bilevel optimization problem, which identifies parameters satisfying two market properties: cost recovery and revenue adequacy. We demonstrate the performance of the proposed approach using a variety of test systems in terms of cost and dispatch approximation errors, distance to feasibility, and generalizability with respect to N-1 network topologies.

4 - Data-driven robust optimal design and control of floating offshore windfarms

Natalie Isenberg, Pacific Northwest National Laboratory, Richland, WA, United States, Bruno Jacob, Jan Drgona, Draguna Vrabie

The unpredictable nature of wind resources presents significant challenges in planning and operating floating offshore wind farms. Efficient wind farm design must account for the uncertainty in wind conditions to prevent substantial efficiency losses and ensure reliable power generation. While many studies have focused on deterministic models and uncertainty quantification for wind farm planning, fewer have addressed optimal design under uncertainty. This work aims to identify robust optimal designs and control policies for floating offshore wind farms that can withstand uncertainty in wind characteristics. We first derive a fully equation-oriented, continuous, nonlinear, and non-convex steady-state model of a wind farm based on parametric wake effect models. A robust optimization methodology is then paired with data-driven approaches to construct uncertainty sets, aiming to find the best wind farm design and control under worst-case wind characteristic scenarios. This is particularly important for floating wind farms, which face greater variability and challenges due to their offshore environments and mobility.

SA32

Summit - 423

Recent Advances in Bilevel Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Erfan Yazdandoost Hamedani, The University of Arizona, Tucson, AZ, United States

Co-Chair: Mahyar Fazlyab, Johns Hopkins University, Baltimore, MD, United States

1 - SimFBO: Towards Simple, Flexible and Communication-efficient Federated Bilevel Learning

Kaiyi Ji, University at Buffalo, Buffalo, NY, United States, Yifan Yang

Federated bilevel optimization (FBO) has shown great potential recently in machine learning and edge computing due to the emerging nested optimization structure in meta-learning, fine-tuning, hyperparameter tuning, etc. However, existing FBO algorithms often involve complicated computations and require multiple sub-loops per iteration, each of which contains a number of communication rounds. In this paper, we propose a simple and flexible FBO framework named SimFBO, which is easy to implement without sub-loops, and includes a generalized server-side aggregation and update for improving communication efficiency. We further propose System-level heterogeneity robust FBO (ShroFBO) as a variant of SimFBO with stronger resilience to heterogeneous local computation. We show that SimFBO and ShroFBO provably achieve a linear convergence speedup with partial client participation and client sampling without replacement, as well as improved sample and communication complexities. Experiments demonstrate the effectiveness of the proposed methods over existing FBO algorithms.

2 - Single-loop Projection-free and Projected Gradient-based Algorithms for Nonconvex-concave Saddle Point Problems with Bilevel Structure

Mohammad Mahdi Ahmadi, University of Arizona, Tucson, AZ, United States, Erfan Yazdandoost Hamedani

In this talk, we investigate a broad class of constrained saddle point problems with a bilevel structure. Here, the upper-level objective function is nonconvex-concave and smooth over compact and convex constraint sets, subject to a strongly convex lower-level objective function. To contribute to the existing literature, we consider a more general setting where the upper-level function is not necessarily strongly concave or linear in the maximization variable. Moreover, we propose a one-sided projection-free method employing a linear minimization oracle. In particular, through regularization and nested approximation techniques, we introduce one-sided projection-free and projected gradient-based algorithms capable of achieving an ε-stationary solution within $O(\varepsilon^{-4})$ and $O(\varepsilon^{-5}\log(1/\varepsilon))$, respectively. When the upper-level objective function is linear in the maximization component, our results improve to $O(\varepsilon^{-3})$ and $O(\varepsilon^{-4})$, respectively.

3 - Prediction-Correction Bilevel Optimization in Continuous Time Models

Nazanin Abolfazli, University of Arizona, Tucson, AZ, United States, Mahyar Fazlyab, Erfan Yazdandoost Hamedani

This Work introduces an advanced continuous-time dynamical system framework designed to address bilevel optimization problems, which are characterized by interdependent decision-making across two levels and are prevalent in disciplines such as engineering, economics, and transportation. The proposed approach leverages continuous-time dynamics to simultaneously resolve upper and lower-level objectives, providing a potent alternative to conventional discrete-time methods.In this study we obtain the convergence rates of our algorithm, both with and without lower-level constraints. Additionally, we conduct extensive numerical experiments to demonstrate the efficacy and superior performance of our method compared to existing solutions.

SA33

Summit - 424

Advances in Data-driven Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Hoda Bidkhori, George Mason University, Fairfax, VA, United States

1 - Robust Quickest Change Detection in Non-Stationary Processes

Yingze Hou, University of Pittsburgh, Pittsburgh, PA, United States

Exactly and asymptotically optimal algorithms are developed for robust detection of changes in non-stationary processes. In non-stationary processes, the distribution of the data after change varies with time. The decision maker does not have access to precise information on the post-change distribution. It is shown that if the post-change non-stationary family has a distribution that is least favorable in a well-defined sense, then the algorithms designed using the least favorable laws are robust optimal. This is the first result where an exactly robust optimal solution is obtained in a non-stationary setting, where the least favorable law is also allowed to be non-stationary. Examples of non-stationary processes encountered in public health monitoring and space and military applications are provided. Our robust algorithms are also applied to real and simulated data to show their effectiveness.

2 - Optimizing Inventory Policies in a Multi-echelon Inventory System Using a Lead Time Approximation

Sharmine Akther Liza, Lehigh University, Bethlehem, PA, United States, Larry Snyder

This study discusses a heuristic for determining an inventory policy for each stage in a general multi-echelon inventory system, which exhibits stochastic lead times because of random stock-outs throughout the system. Stochastic lead times pose challenges not only for general multi-echelon inventory systems but even for simpler, serial systems. Dynamic programming (DP) has mostly been used in single-stage or serial inventory systems considering deterministic lead times. But with stochastic lead times it becomes impossible to solve because of the larger state space. Our proposed heuristic addresses this problem by approximating the lead time distribution based on stock levels of each stage in the inventory system. Ultimately, the DP algorithm uses this approximate lead time distribution to approximate the demand distribution, which facilitates optimizing the optimal inventory policy for each stage.

3 - Effective Scenarios in Distributionally Robust Optimization with Wasserstein Distance

Guzin Bayraksan, The Ohio State University, Columbus, OH, United States, Chennan Zhou

We investigate *effective scenarios* in Distributionally Robust Optimization (DRO) problems where the ambiguity set of distributions are defined on a finite number of realizations (also called scenarios) of the uncertain parameters using the Wasserstein distance. *Effective scenarios* are critical scenarios in DRO in the sense that their removal from the support of the considered distributions alters the optimal value. *Ineffective scenarios* are those whose removal do not alter the optimal value. We first link the effectiveness of a scenario to its worst-case probability being always positive or uniquely zero under a general ambiguity set. We then narrow our focus to DROs with ambiguity sets formed via the Wasserstein distance (DRO-W) and provide easy-to-check sufficient conditions to identify the effectiveness of scenarios for this class of problems. When the Wasserstein distance is equivalent to the total variation distance (i.e., when the transportation cost between scenarios is zero if they are the same and one if they are different), the easy-to-check conditions for DRO-W presented in this talk recover the ones presented in the literature for DRO formed via the total variation distance as a special case. The numerical findings highlight the relationship between scenario effectiveness and the attributes of the transportation cost between scenarios that constitute the Wasserstein distance, revealing useful insights.

4 - Data-Driven Optimization in Partially Observable Environments

Hoda Bidkhori, George Mason University, Fairfax, VA, United States

Missing data is a common issue for many practical data-driven stochastic programming problems. The state-of-the-art approaches first estimate the missing data values and then separately solve the corresponding stochastic programming. Accurate estimation of missing values is typically inaccessible as it requires enormous data and sophisticated statistical methods. Therefore, this paper proposes an integrated approach, a distributionally robust optimization (DRO) framework, that simultaneously tackles the missing data problem and data-driven stochastic optimization by hedging against the uncertainties of the missing values. This paper adds to the DRO literature by considering the practical scenario where the data can be incomplete and partially observable; it particularly focuses on data distributions with finite support.

We construct several classes of ambiguity sets for our DRO model utilizing the incomplete data sets, maximum likelihood estimation method, and different metrics. We prove the statistical consistency and finite sample guarantees of the corresponding models and provide tractable reformulations of our model for different scenarios. We perform computational studies on the multi-item inventory control problem and portfolio optimization using synthetic and real-world data. We validate that our method outperforms the traditional estimate-then-optimized approaches

SA34

Summit - 425

New Advances in Large-Scale Non-Linear Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Miaolan Xie, Cornell University, Ithaca, NY, United States

1 - Sample Sizing for Function Estimation in Stochastic Derivative-Free Optimization

Luis Nunes Vicente, Lehigh University, Bethlehem, PA, United States

We introduce a new tail bound condition for function estimation which allows for an improvement in the required sample sizing for derivative-free optimization of the order delta^{-2q} (non-correlated errors) and delta^{2-2q} (correlated errors), where delta is the step size or trust-region radius and the error is assumed to have a momentum of order q/(1-q). We also introduce simple stochastic non-smooth direct-search and trust-region methods, sharing the following principles: generate a direction; generate the new iterate by either search along the direction or by solving a trust-region subproblem (where the linear term is the direction); use a sufficient decrease acceptance test to decide if the trial point can be accepted or not. The methods are shown to globally converge almost surely under the new tail bound condition. We will also report more recent results on sequential sampling. Joint work with F. Rinaldi and D. Zeffiro.

2 - Searching for Optimal Per-Coordinate Step-Sizes with Multidimensional Backtracking

Frederik Kunstner, University of British Columbia, Vancouver, BC, Canada, Victor Sanches Portella, Mark Schmidt, Nicholas Harvey

The backtracking line-search is an effective technique to automatically tune the step-size in smooth optimization. It guarantees similar performance to using the theoretically optimal step-size. Many approaches have been developed to instead tune per-coordinate step-sizes, also known as diagonal preconditioners, but none of the existing methods are provably competitive with the optimal per-coordinate step-sizes. We propose multidimensional backtracking, an extension of the backtracking line-search to find good diagonal preconditioners for smooth convex problems. Our key insight is that the gradient with respect to the step-sizes, also known as hyper-gradients, yields separating hyperplanes that let us search for good preconditioners using cutting-plane methods. As black-box cutting-plane approaches like the ellipsoid method are computationally prohibitive, we develop an efficient algorithm tailored to our setting. Multidimensional backtracking is provably competitive with the best diagonal preconditioner and requires no manual tuning.

3 - First-Order Methods for Bilevel Optimization

Zhaosong Lu, University of Minnesota, Minneapolis, MN, United States, Sanyou Mei

Bilevel optimization, also known as two-level optimization, is an important branch within mathematical optimization. It has found applications across various domains, including economics, logistics, supply chain, transportation, engineering design, and machine learning. In this talk, we will present first-order methods for solving a class of bilevel optimization problems using either single or sequential minimax optimization schemes. We will also discuss the first-order operation complexity of these methods and present preliminary numerical results to illustrate their performance.

4 - On Existence of Solutions to Non-Convex Minimization Problems

Rohan Rele, ASU, Tempe, AZ, United States, Angelia Nedich

We describe sufficient conditions for existence of a solution to a general minimization problem of a closed, proper objective function on nonempty, closed feasible set. We introduce a novel relaxation of coercivity and exploit the asymptotic geometry of unbounded feasible regions to unify previously known existence results of Frank-Wolfe-type. We show that our conditions apply to a broad class of objective functions and feasible regions in both convex and non-convex settings.

SA35

Summit - 427

Improving Societal Outcomes with Logistical Innovations

Invited Session

TSL: Air Transportation

Chair: Amro El-Adle, Strome College of Business, Old Dominion University, Norfolk, VA, United States

1 - Fairness in Drone Delivery Flight Paths

Amro El-Adle, Strome College of Business, Old Dominion University, Norfolk, VA, United States, Ahmed Ghoniem, Blake Steiner

We investigate the impact of noise pollution on flight path planning for last-mile delivery via drone. Over an extended planning horizon, the proposed mathematical model optimizes the selection of flight paths used to mitigate noise pollution for residents in the path of the drones. Our results reveal that a set of operational strategies may greatly mitigate not only the overall level of noise pollution, but may also result in a more equitable distribution of the noise from drone flights.

2 - A decomposition approach for parcel delivery by vehicle and drone over ordered neighborhoods

Ahmed Ghoniem, Isenberg School of Management, UMass Amherst, Amherst, MA, United States, Semih BOZ, Amro El-Adle

We examine a multi-vehicle last-mile delivery problem with companion drones over pre-ordered residential neighborhoods. The problem is formulated as a mixed-integer program that is enhanced with valid inequalities. Alternatively a decomposition technique is proposed and tested over problem instances based on the town of Amherst, MA. Computational results are discussed.

3 - Parcel Consolidation In Last Mile Delivery by Vehicle and Drone

Semih BOZ, Isenberg School of Management, UMass Amherst, Amherst, MA, United States, Ahmed Ghoniem, Amro El-Adle

We study parcel consolidation within the context of last-mile delivery, employing both a vehicle and a companion drone to efficiently serve customers across multiple time periods. We present a mixed-integer program enhanced with RLT based valid inequalities and an optimization-based heuristic to address large-scale instances. The efficacy of our methodology is assessed through computational experiments on randomly generated instances, focusing on a residential network in Amherst, Massachusetts. Additionally, a sensitivity analysis is conducted with respect to selected parameters.

SA36

Summit - 428

Innovative Strategies in Freight and Logistics: From Air Mobility to Electric Trucks

Invited Session

TSL: Freight Transportation

Chair: Ruiting Wang, University of California, Berkeley, Berkeley, CA, United States

1 - Optimal Dispatching in Two-Sided Spatial Queues

Ang Xu, University of California, Berkeley, Berkeley, CA, United States, Chiwei Yan

We study an optimal dispatching problem in two-sided spatial queues motivated by ride-hailing platforms. The platform aims to maximize the system throughput (or revenue) penalized by waiting cost by controlling the matching process through a Markov Decision Process framework. Under some mild conditions, we give closed-form state-dependent dispatching policies. In the general case, we develop an efficient dynamic programming algorithm for finding optimal policy within a sub-class of dispatching policies.

2 - A TIME-Invariant Network Flow Model for Ride-Pooling in Mobility-on-Demand Systems

Fabio Paparella, Eindhoven University of Technology , Eindhoven, Netherlands

This paper presents a framework to incorporate ride-pooling from a mesoscopic point of view, within time-invariant network flow models of Mobility-on-Demand systems. The resulting problem structure remains identical to a standard network flow model, a linear problem, which can be solved in polynomial time. In order to compute the ride-pooling assignment, which is the matching between two or more users so that they can be pooled together, we devise a polynomial-time knapsack-like algorithm that is optimal w.r.t. the minimum vehicle travel time with users onboard. Finally, we conduct two case studies of Sioux Falls and Manhattan, where we validate our models against state-of-the-art results, and we quantitatively highlight the effects that maximum waiting time and maximum delay thresholds
have on the vehicle hours traveled, overall pooled rides and actual delay experienced. Last, we show that allowing for four people ride-pooling can significantly boost the performance of the system.

3 - Integrating Unmanned Aircraft Vehicles into Grid

Jiaman Wu, UC Berkeley, Berkeley, CA, United States

The rapid growth of e-commerce, coupled with the urgent need for decarbonization and advancements in technology, is revolutionizing urban package delivery. Among various innovations being considered, delivery by Unmanned Aircraft Vehicles (UAVs) is increasingly seen as a component with great potential for future urban freight movement, as it offers rapid and point-to-point deliveries. With a potentially significant demand for UAV delivery in the future, there is a growing need to efficiently manage UAV operations in urban environments. Since UAV operations rely heavily on battery charging and fast-charging capabilities, optimizing their charging schedules is crucial for integrating UAVs into the grid effectively. This study aims to improve the balance of loads on the electric grid and minimize the impact on power grid loads, voltage, frequency, and power losses by controlling UAVs' charging schedules.

4 - Joint Optimization of Charging Infrastructure Placement and Operational Schedules for a Fleet of Battery Electric Trucks

Juan Pablo Bertucci, Eindhoven University of Technology, Eindhoven, Netherlands, Theo Hofman, Mauro Salazar

In the immediate advent of zero-emission zones, fleet operators are transitioning to electric fleets. To maintain their current operations, a clear understanding of the required charging infrastructure and its relationship to existing power grid limitations and fleet schedules is essential.

In this context, we introduce an optimization framework for designing the charging infrastructure and creating routing and charging schedules for a logistic distribution network. We validate our results using agent-based simulations.

Specifically, we start by formulating the joint infrastructure design and operational scheduling problem as a mixed-integer linear program. Next, we develop an agent-based model that acts as a virtual experimental replica to evaluate various design and operational solutions.

Finally, we conduct experiments comparing rule-based approaches to optimized design and operation strategies across key operational metrics, as well as power requirements. Our case study focuses on the logistics network in The Netherlands.

The results reveal that distributed vehicle charging rules mitigate the impact of consumption uncertainties, thereby enhancing logistics reliability. However, central coordination leads to more cost-efficient solutions (16.1% to 35.4% energy cost reductions) and less stringent power requirements (22.8% to 47.9% less time operating at peak consumption).

Explainable optimization and simulation frameworks, incorporating real-case reliability metrics, play a crucial role in improving decision-making and facilitating a smoother transition to battery electric trucks.

SA37

Summit - 429

Semiconductor Industry and Batch Production

Contributed Session

Contributed

Chair: julien Maurice

1 - Reinforcement Learning-based Facility Layout Redesign to Improve Logistics Efficiency on a Semiconductor FAB

Ha Neul Kim, Korea University, Seoul, Korea, Republic of, Junghoon Kim, Subin Oh, Sangmin Lee, Taesu Cheong

This research aims to improve logistics efficiency through M:N exchange of facilities based on the existing facility layout in a semiconductor FAB. To produce a semiconductor, multiple processes must be performed along the processing sequence. After each process is completed, work in progress (WIP) is transported to the next process facility, thus the flow of logistics has a significant effect on productivity. However, the facility layout within a FAB is typically oriented towards maximizing production capacity within a limited space, and optimizing the flow of logistics is often disregarded. Our goal is to minimize the sum of the distances between any two facilities, each belonging to consecutive processes, through the exchange of interchangeable facility sets based on the existing layout. This problem is characterized by changes in distances between exchanged facilities and remaining facilities after any exchange occurs. Therefore, we propose a dynamic programming algorithm, and in addition, an attention network-based deep reinforcement learning model is proposed to consider the computational complexity. Through experiments, we compare the performance between our proposed reinforcement learning model and other heuristic algorithms.

2 - Continual learning-based remaining useful life prediction of machining tools under varying operating conditions

Gyeongho Kim, Ulsan National Institute of Science and Technology, Ulsan, Korea, Republic of, Yun Seok Kang, Sang Min Yang, Jae Gyeong Choi, Gahyun Hwang, Hyung Wook Park, Sunghoon Lim

Data-driven prediction of the remaining useful life (RUL) of equipment and assets has become one of the most important tasks in manufacturing. Accurate RUL prediction not only helps prevent unexpected failures but also enables maximal utilization of available life, thus contributing to improved process efficiency. However, in practice, the use of multiple operating conditions that vary by time impedes performing an efficient data-driven RUL prediction. Unlike conventional supervised learning setups, varying operating conditions generate heterogeneous data of which the generating distribution also is time-varying. Thus, existing approaches cannot be effectively applied due to linearly increasing modeling and memory costs. One of the domains that suffer from this issue is machining, where accurate RUL prediction of cutting tools is extremely important for productivity. Considering the realistic circumstances where operating conditions vary over time, this work proposes Fisher-informed continual learning (FICL), which enables an efficient tool RUL prediction by using a single model that adaptively learns as condition changes without the need to store previous data and models. In particular, Fisher information is used to improve generalization via sharpness-aware minimization and to transfer informative knowledge between operating conditions through regularization-based continual learning. Experiments are performed using datasets collected from real-world machining processes under five distinct operating conditions. The results prove the efficacy of the proposed FICL, indicating its superior RUL prediction performance to existing data-driven approaches for all operating conditions. In addition, the proposed FICL has shown the least catastrophic forgetting, implying that it effectively retains informative knowledge from previous operating conditions.

3 - The concept of Hyperconnected Mobile Supply Chains to address the distributed production near customers problem.

julien Maurice, H. Milton Stewart School of Industrial and Systems Engineering,Georgia Institute of Technology, Atlanta, GA, United States, Simon Kwon, Benoit Montreuil

In a world where consumers demand high-quality products at minimum cost and with rapid delivery, traditional production and delivery models are challenged, especially when considering large-sized products such as modular building units. As a solution, the concept of Hyperconnected Mobile Production was developed by Montreuil and Marcotte (2016). This strategy leverages the concepts of modularization, standardization, interfaces, and protocols of the Physical Internet, to allow businesses to dynamically expand and contract their production capacity in a network of open certified facilities. In their paper, they limit to a single-tier setting. However, supply chains are oftentimes multi-tier multi-party systems that require interaction between multiple production networks. The gap leads us to introduce the concept of Hyperconnected Mobile Supply Chains, extending from single-tier multi-party production networks to multi-tier multi-party supply chains, linking multiple mobile production networks with plug-and-play modular mobile production units across the supply tiers and enabling creation of flexible, responsive, and environmentally friendly supply chains.

Several criteria impact the pool of facilities to consider, such as cost, environmental, or societal objectives. We propose a multi-objective mixed integer programming formulation for the strategic network design of hyperconnected mobile supply chains to serve a dynamic set of customers across a wide territory, leveraging existing open production facilities and proposing location and capacity of supplementary facilities. We propose a solution approach leveraging the epsilon-constraint method. We present experimental results from an industrial case to assess the computational performance and the economic and environmental benefits brought by Physical Internet enabled Hyperconnected Mobile Supply Chains.

SA38

Summit - 430

Emerging Delivery Technology

Invited Session

Aviation Applications

Chair: Yuqian Dong, Walmart, Seattle, WA, 98119, United States

Co-Chair: Fanruiqi Zeng, Georgia Institute of Technology, Atlanta, GA, United States

1 - Deploying Next Generation Technologies for Aerial Urban and Rural Delivery

Mark Moore, Whisper Aero, Crossville, TN, United States

Aerial delivery is an emerging direct delivery option for smaller packages that leverages flying drone technologies. The operational aspects for these deliveries are highly complex, as significant environmental constraints are present, especially in densely populated urban areas for the final 100' distance. Varied approaches are being operationally tested, including use of parachutes, winches, building attachment, portals, car rooftop, and ground landing - each facing their own obstacles towards achieving high reliability, minimum risk, and maximum community acceptance. The author's involvement in the UberEats drone delivery project provided valuable insight across these options that included test deliveries with partners. A common obstacle across all approaches is drone noise as one of the most significant obstacles for community acceptance. Another community acceptance obstacle is the perceived safety of operations, particularly in the event of a component failure. New technology approaches offer methods to mitigate both of these concerns. Ensuring the merging of new technologies, opportunities and operational approaches can achieve not only reliability, risks and community acceptance - but also meet aggressive platform and delivery costs, which is critical for drone delivery to reach scaled operations. The presentation will focus on an overview of the drone delivery opportunity, existing developers and their specific approaches, a requirement perspective for what's needed, and Whisper Aero's electric electric ducted fan technology that eliminates exposed blades to achieve high efficiency and ultra-quiet operations.

2 - Noise Analysis for Drone Delivery Operations

Raphael Gautier, Georgia Institute of Technology, Atlanta, GA, United States, Holger Pfaender, Dimitri Mavris

The unmanned aircraft system (UAS) market, which includes drone package delivery, is expected to grow rapidly in the coming years, possibly leading to large-scale operations whose environmental impact will have to be assessed to be authorized. However, UAM operations bring unique requirements that are not addressed by today’s noise assessment capabilities, which focus on commercial aviation operations around airports during the departure and arrival phases of flights. UAM operations are expected to be denser than current general or commercial aviation operations, requiring the ability to analyze many flights simultaneously. UAM vehicles are expected to be relatively small and quiet, leading to more localized noise footprints than traditional aviation operations, calling for higher spatial resolution of analyses. Finally, UAM operations are expected to be stochastic in nature with day-to-day variations stemming from demand. The objective of this research is to develop a novel noise estimation method that supports the computation of noise resulting from the stochastic operation of UAS with irregular locations and operations in large numbers. To achieve this, we propose a multi-step approach: 1) demand estimation, leveraging publicly available US census data, 2) planning of operations, accounting for diverse concepts of operations and the presence of multiple operators, 3) path planning, to estimate the actual trajectory followed by vehicles, 4) noise propagation, where source noise from operating vehicles is propagated to a dense grid of virtual receivers. We demonstrate the proposed approach on a notional use case featuring multiple operators over the Atlanta urban area.

3 - Towards safe and fluent interactions for autonomous systems

Karen Leung, University of Washington, Seattle, WA, United States

Autonomous robots are becoming increasingly prevalent in our daily existence, from navigating our roads and sky, assisting in households and warehouses, conducting daring search and rescue missions, and even exploring the frontiers of space. Yet building robots that can safely and fluently interact with humans in a trusted manner remains an elusive task. Ensuring robots can keep a sufficiently safe distance from humans is at odds with fluent interactions, yet humans are remarkable at seamlessly avoiding collision in crowded settings. In this talk, we will study how humans engage in safe and fluent multi-agent interactions and how this can be applied to robot decision-making and control. With a few relatively simple tweaks to a general trajectory optimization formulation, we can induce legible and proactive robot behaviors that can lead to prosocial interactions. This talk aims to rethink how safety and fluency can be more compatible with one another, leading to more natural and trusted autonomous operations.

4 - Walmart Drone Delivery

Coline Ramee, Walmart, Sunnyvale, CA, United States, Yuqian Dong, Xinyue Peng, Precia Carraway

With 90% of the US population located within 10 miles of one of its stores, Walmart is uniquely positioned to provide drone delivery service to its customers, making a large assortment of products available in less than 30 minutes without leaving the comforts of home.

Initial pilots have proven the feasibility, interest and acceptance from the local communities. Walmart is currently expanding aggressively around Dallas and plans to cover 75% of single-family homes in this market by early 2025. This will be used to prove the economic viability of the drone delivery service and observe customer behavior.

To identify the best markets to launch drone deliveries, feasibility, demand, and coverage were considered.

Today, drones can only operate in fair weather and are restricted to operate away from airports, removing 25% of potential locations.

A granular demand model, accounting for customers affinity for drone delivery based on their distance to stores, was developed. Current customer orders were used to estimate the basket value that could be expected for drone orders based on the payload weight and dimensions.

Store selection was optimized based on operational constraints to maximize satisfied demand while minimizing operating costs. Other Walmart initiatives, such as increasing fulfillment automation in the stores, are being leveraged during the selection to reduce the cost to serve.

This approach allowed us to define a coherent strategy for drone in a rapidly evolving technological and regulatory landscape. The assumptions and results of the model will be validated in the coming year in Dallas.

SA40

Summit - 432

Behavior in Queueing and Service Systems

Invited Session

Behavioral Operations Management

Chair: Blair Flicker, University of South Carolina, Columbia, SC, United States

1 - Give Me a Choice! a Field Experiment on Task Choice Enabled by Wearables

Daniel Kwasnitschka, University of South Carolina, Columbia, SC, United States, Henrik Franke, Torbjorn Netland

Considering human motivation, is it more effective to delegate work tasks to shop floor workers or let them choose among a set of open tasks? This paper reports the results of a field experiment conducted with a manufacturer that uses wearable devices to distribute shop floor tasks. We show that a system that allows workers to choose their next task outperforms a task delegation system in productivity but does not affect work speed and decreases workers' task responsiveness. We measure the effect of giving workers the choice to choose the next task using a Difference-in-Difference approach in a 13-week-long field experiment. We collect data from 65,348 machine status reports and 49,050 work tasks completed in two manufacturing plants in Germany and Italy. Our results indicate a productivity increase of 6.9%, which corresponds to an estimated annual savings potential of USD 98,168. In summary, our results both align with and challenge aspects of organizational theories on human motivation in work settings. This offers interesting implications for production managers considering the implementation of digital task-distribution systems.

2 - Experienced and Prospective Wait in Queues: a Behavioral Investigation

Jing Luo, University of Science and Technology Beijing, Beijing, China, People's Republic of, Leon Valdes, Sera Linardi

Problem definition: The relative contribution of the length, speed, and wait time of a queue to people’s disutility is not well understood. Empirical studies show deviations from hyper-rationality, but the impact of each factor—separate from one another and differentiating their effects in the experienced (elapsed) vs. prospective (remaining) queue—is unclear. This problem is compounded by the fact that the continuous cost of completing a queue, V , is often studied only indirectly through binary, censored decisions like balking or reneging. Methodology/results: Using a series of controlled experiments, we introduce the use of the Becker-DeGroot-Marschak (BDM) method to directly measure V in deterministic, visible queues. With it, we study the effects of our key parameters of interest concurrently, hence avoiding omitted- variable biases. We find that customers consider the length and speed of the queue ahead of them additively, not multiplicatively as they would if they were responding proportionally to total wait time. Regarding the experienced queue, we observe that in most cases, it does not affect completion costs. The only exception is in the slowest of our queues, and only when participants have made significant progress, when we find signs of commitment through lower completion costs. Managerial Implications: Organizations regularly offer customers paid options to shorten their wait time, present them with alternatives to joining a line, or use incentives to manage congestion. Our work provides them with a granular metric of completion costs, measured in monetary terms, as well as an understanding of their dependence on queuing factors.

3 - How a Wait's Presentation Affects Its Perceived Cost

Blair Flicker, University of South Carolina, Columbia, SC, United States, Charlie Hannigan

Physical queues are being replaced by virtual waiting systems. Managers must now decide how to communicate wait characteristics to customers, and we show that their choice of presentation scheme can substantially influence behavior. This paper describes two controlled experiments. The first focuses on customers’ accuracy in forecasting the duration of hypothetical waits while the second studies customers’ preferences over real waits. In both, we study two wait presentation formats: Clocks and Queues. Clocks (e.g., apps displaying expected waiting times) excel at clearly communicating total wait time to customers, but they emphasize wait duration which increases people’s perceived disutility of the wait. Queues (e.g., wait lists showing a position in line), in contrast, make it difficult for customers to derive total wait time. When facing Queues, people’s wait time forecasts are systematically too short which makes Queue-framed waits appear more attractive than Clock-framed analogs. We that estimate the perceived cost waiting is twice as high for Clocks vs. Queues. We also find that people are risk averse with their time and that this has a large influence on disutility. Finally, we organize our results into a parsimonious behavioral utility function to add more nuance to queueing models.

4 - Multitasking over TIME: the Time-Dependent Effects of Multitasking

Bob Batt, Wisconsin School of Business, UW-Madison, Madison, WI, United States, Santiago Gallino

This paper explores agent multitasking over time in customer service environments, particularly within livechat settings. Multitasking, a ubiquitous occurrence in various professional sectors, is typically assumed to have a constant effect conditional on the multitasking level. We go beyond this conventional approach by examining the temporal dynamics of multitasking and their effects on performance. Through a collaboration with a third-party customer contact center, we analyze livechat interactions across different retail brands, focusing on both multitasking and multibranding, where agents manage tasks for multiple brands simultaneously. Our findings reveal that handling multiple chats simultaneously increases the average chat duration, with diminishing marginal increases as the number of concurrent tasks grows. We explore the implications of these insights by proposing a “pile-on” heuristic, which assigns new tasks to agents already engaged in multitasking, rather than evenly distributing tasks across agents. This method can lead to a reduction in average chat duration and increased agent idle time. Our study contributes to the operations management literature by illustrating the dynamic nature of multitasking, challenging traditional perceptions and suggesting that the duration and intensity of multitasking play crucial roles in shaping operational efficiency, and we show how work assignment strategies can influence operational outcomes.

SA41

Summit - 433

Learning in Revenue Management

Invited Session

Applied Probability Society

Chair: Rui Gao, University of Texas at Austin, Austin, TX, United States

Co-Chair: Zhi Wang, University of Texas at Austin, Austin, TX, United States

1 - A Unified Analysis for Assortment Planning with Marginal Distributions

Xiaobo Li, National University of Singapore, Singapore, Singapore, Zeyu Sun, Selin Ahipasaoglu

In this paper, we examine assortment planning using the marginal distribution model (MDM), a semiparametric choice model that relies solely on information about the marginal noise in alternative utilities without assuming noise term independence. While previous literature has established the multinomial logit (MNL) model as part of the MDM framework, we extend this to include other multi-purchase choice models like the multiple-discrete-choice (MDC) and threshold utility model (TUM), despite MDM not explicitly addressing multi-purchase behavior. Within the MDM framework, we establish a condition for optimal strictly profit-nested assortments for assortment problems. Although NP-hard, we demonstrate that the best strictly profit-nested assortment provides a 1/2-approximate solution for all MDMs, with an example illustrating this bound's tightness. These findings either expand or refine previous insights on assortment optimization under MNL, MDC, and TUM. We also show how these bounds extend naturally to mixtures of MDM models. Furthermore, empirical experiments on a real-world dataset with multi-purchase behavior compare the predictive capabilities of various MDM choice models. Our results indicate that, beyond MDC, several other MDMs exhibit promising predictive power in multi-purchase scenarios, emphasizing the benefits of analyzing multi-purchase behavior within the MDM framework.

2 - When Online Resource Allocation Meets Watermarking LLMs

Xiaocheng Li, Imperial College Business School, London, United Kingdom, Hanzhao Wang, Shang Liu, Huaiyang Zhong, Zhongze Cai

We study the problem of watermarking LLMs, where a watermark for LLMs is a hidden pattern injected to the generated text that is imperceptible to humans but can be algorithmically detected as generated by AI. We consider the trade-off between model distortion and detection ability, and formulate the watermarking procedure as a constrained optimization problem based on the green-red algorithm of Kirchenbauer et al. (2023). We show that the optimal solution to the optimization problem enjoys a nice analytical property. This analytical structure renders the formulation fall under the framework of online optimization with constraints where the existing online resource allocation algorithms in OR/MS literature can be naturally applied. These developments also shed light on the properness of model distortion metrics and the repetition phenomenon for the watermarking algorithms.

3 - Consider OR Choose? the Role of Consideration Sets in Discrete Choice Modeling

Yi-Chun Akchen, University College London, London, United Kingdom, Dmitry Mitrofanov

Consideration sets play an crucial role in discrete choice modeling, where a large class of non-parametric models can be understood as a two-stage decision making process. In this process, customers first form consideration sets and then follow a choice mechanism, such as a ranking system or a decision tree, to navigate their way towards a purchase decision. Rather than delving into a more sophisticated second-stage choice mechanism, this paper takes a step back and focuses on the fundamental properties of the decision making process based on the consideration sets. We consider a class of nonparametric choice models that is only specified by a distribution over consideration sets and has a bounded rationality interpretation. We call it the *consideration set model*. Intriguingly, we show that this class of choice models can be characterized by the axiom of symmetric demand cannibalization and enable full statistical identification. We further consider the model's downstream assortment planning as an application. We first provide an exact characterization of the optimal assortment and show that it is revenue-ordered on the blocks defined with respect to the consideration sets. Despite the existence of the exact structure, we prove that the optimal assortment is NP-hard to approximate. This result shows that the existence of consideration sets inevitably cause inapproximability in assortment planning, despite that the consideration set model does not involve a sophisticated second-stage mechanism. Finally, using a real-world dataset, we show that the consideration set model offers comparable prediction power to other advanced choice models.

4 - Learning Mixed Logit Choice Model with Neural Network: Global Convergence and Generalization Bound

Zhi Wang, University of Texas at Austin, Austin, TX, United States, Rui Gao

The mixed logit model is an expressive class of discrete choice models that have been widely studied in operations, marketing, and econometrics. However, estimating the mixing distribution presents significant challenges. Existing nonparametric methods often lack global convergence guarantees (such as the Expectation-Maximization algorithm), or necessitate non-convex optimization subroutines to secure global convergence (such as the Frank-Wolfe algorithm).

This study explores the efficacy of a one-hidden-layer neural network model to approximate the mixing distribution. Leveraging recent advances in the mean-field analysis of neural networks, we demonstrate that the gradient descent algorithm can identify the globally optimal estimator, subject to an approximation error. This error is inversely proportional to the neural network's width and becomes negligible in over-parameterized models. Furthermore, we prove that over-parameterization does not compromise generalization on unseen data, by deriving width-independent generalization bounds for finite-width networks. Using both synthetic and real datasets, we validate our algorithm's superior in-sample and out-of-sample performance against existing benchmarks.

Our findings underscore the potential of even shallow neural networks, armed with efficient training algorithms, to effectively learn complex choice models. The resultant model not only boasts high predictive accuracy but also exhibits guaranteed generalization capabilities.

SA42

Summit - 434

Yunbei Xu and Ayush Sekhari session

Invited Session

Applied Probability Society

Chair: Ayush Sekhari, MIT, Cambridge, United States

Co-Chair: Yunbei Xu, National University of Singapore, Singapore

1 - Computationally Efficient Reinforcement Learning with Linear Bellman Completeness

Noah Golowich, MIT, Cambridge, MA, United States, Ankur Moitra

One of the most natural approaches to reinforcement learning (RL) with function approximation is value iteration, which inductively generates approximations to the optimal value function by solving a sequence of regression problems. To ensure the success of value iteration, it is typically assumed that Bellman completeness holds, which ensures that these regression problems are well-specified. We study the problem of learning an optimal policy under Bellman completeness in the online and offline models of RL with linear function approximation:

- In the online setting, while statistically efficient algorithms are known under Bellman completeness, these algorithms all rely on the principle of global optimism which requires solving a nonconvex optimization problem. In particular, it has remained open as to whether computationally efficient algorithms exist. We give the first polynomial-time algorithm for online RL under linear Bellman completeness when the number of actions is any constant.

- In the offline setting, we give the first algorithm which finds a near-optimal policy under a single-policy coverage condition. Our algorithm is also computationally efficient and generalizes to the setting with small inherent Bellman error, for which the suboptimality of our output policy scales with the square root of the Bellman error. We also prove a matching lower bound, showing that no algorithm can achieve better than square-root dependence on the Bellman error. This lower bound stands in contrast to many other settings in reinforcement learning with misspecification, where one can typically obtain performance that degrades linearly with the misspecification error.

2 - Offline Reinforcement Learning: Role Of State Aggregation and Trajectory Data

Zeyu Jia, Massachusetts Institute of Technology, Cambridge, MA, United States

We revisit the problem of offline reinforcement learning with value function realizability but without Bellman completeness. Previous work by Xie and Jiang (2021) and Foster et al. (2022) left open the question whether a bounded concentrability coefficient along with trajectory-based offline data admits a polynomial sample complexity. In this work, we provide a negative answer to this question for the task of offline policy evaluation. In addition to addressing this question, we provide a rather complete picture for offline policy evaluation with only value function realizability. Our primary findings are threefold: 1) The sample complexity of offline policy evaluation is governed by the concentrability coefficient in an aggregated Markov Transition Model jointly determined by the function class and the offline data distribution, rather than that in the original MDP. This unifies and generalizes the ideas of Xie and Jiang (2021) and Foster et al. (2022), 2) The concentrability coefficient in the aggregated Markov Transition Model may grow exponentially with the horizon length, even when the concentrability coefficient in the original MDP is small and the offline data is admissible (i.e., the data distribution equals the occupancy measure of some policy), 3) Under value function realizability, there is a generic reduction that can convert any hard instance with admissible data to a hard instance with trajectory data, implying that trajectory data offers no extra benefits over admissible data. These three pieces jointly resolve the open problem, though each of them could be of independent interest.

3 - Selective Sampling and Imitation Learning via Online Regression, and Implications for Unlearning

Ayush Sekhari, MIT, Cambridge, MA, United States

We consider the problem of Imitation Learning (IL) by actively querying noisy expert for feedback. While imitation learning has been empirically successful, much of prior work assumes access to noiseless expert feedback which is not practical in many applications. In fact, when one only has access to noisy expert feedback, algorithms that rely on purely offline data (non-interactive IL) can be shown to need a prohibitively large number of samples to be successful. In contrast, in this work, we provide an interactive algorithm for IL that uses selective sampling to actively query the noisy expert for feedback. Our contributions are twofold: First, we provide a new selective sampling algorithm that works with general function classes and multiple actions, and obtains the best-known bounds for the regret and the number of queries. Next, we extend this analysis to the problem of IL with noisy expert feedback and provide a new IL algorithm that makes limited queries. I will conclude by discussing the implications of selective sampling for the problem of machine unlearning with limited memory.

4 - Instance-Dependent Complexity of Contextual Bandits and Reinforcement Learning: A Disagreement-Based Perspective

Yunzong Xu, Microsoft Reseach, New York, NY, United States

We introduce a family of complexity measures that are both sufficient and necessary to obtain instance-dependent regret bounds in contextual bandits. We then introduce new oracle-efficient algorithms which adapt to the gap whenever possible, while also attaining the minimax rate in the worst case. Finally, we provide structural results that tie together a number of complexity measures previously proposed throughout contextual bandits, reinforcement learning, and active learning and elucidate their role in determining the optimal instance-dependent regret. Additionally, we compare our results to some more recent work on instance-dependent complexity measures.

5 - Near-Optimal Pure Exploration in Matrix Games: A Generalization of Stochastic Bandits & Dueling Bandits

Kevin Jamieson, University of Washington, Seattle, WA, United States

We study the sample complexity of identifying the pure strategy Nash equilibrium (PSNE) in a two-player zero-sum matrix game with noise. Formally, we are given a stochastic model where any learner can sample an entry (i,j) of the input matrix A ∈ [−1,1]^n×m and observe Ai,j + η, where η is a zero-mean 1-sub-Gaussian noise. The aim of the learner is to identify the PSNE of A, whenever it exists, with high probability while taking as few samples as possible. Zhou et al. [21] present an instance-dependent sample complexity lower bound that depends only on the entries in the row and column where the PSNE lies. We design a near-optimal algorithm whose sample complexity matches the lower bound, up to log factors. The problem of identifying the PSNE also generalizes the problem of pure exploration in stochastic multi-armed bandits and dueling bandits, and our result matches the optimal bounds, up to log factors, in both settings.

SA43

Summit - 435

Algorithmic Accountability

Invited Session

Applied Probability Society

Chair: Sarah Cen, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Machine Unlearning via Simulated Oracle Matching

Seth Neel, Harvard, Boston, MA, United States

Despite increasing interest in machine unlearning, recent work

shows that under strong evaluations, existing techniques largely fail to
unlearn in non-convex settings.
In this paper, we introduce a new technique for machine unlearning in such settings.
Key to our method is a reduction from the problem of machine unlearning
to that of data attribution.
In particular,
we show theoretically (in an underdetermined regression setting)
and empirically (in a standard deep learning setting)
that given access to the outputs of a perfectly unlearned model
(i.e., a model trained from scratch on the non-unlearned data),
we can quickly fine-tune an existing model on these predictions
and match the target model predictions out-of-sample.
Meanwhile, predicting such ``oracle''
outputs is precisely the goal of a recent line of
work in data attribution called {\em datamodeling}.
Combining these two insights yields an end-to-end unlearning
algorithm in which one first {\em predicts the output}
of a model re-trained from scratch, then {\em fine-tunes}
an existing model to match these predicted outputs.
Across different types and sizes of forget sets on standard classification tasks,
we show that this two-stage
algorithm results in strong unlearning performance being close to indistinguishable from the fully-retrained ``oracle'' model in some cases.
As an added benefit, our reduction means that future improvements
to data attribution---whether in accuracy or efficiency---yield better unlearning algorithms.

2 - Black-Box Counterfactual Auditing Using LLMs

Chara Podimata, MIT, Boston, MA, United States, Sarah Cen, Andrew Ilyas

Auditing of social media platforms has become increasingly important, given the impact that these platforms have in shaping beliefs and opinions. For all its importance, auditing remains hard to implement properly and rather costly; an auditor needs to find a lot of users that would be willing to give their data and even in that case, it's unclear how to create proper counterfactuals. In this work, we present a new framework for doing black-box and counterfactual auditing using Large Language Models (LLMs). Our approach is "black-box" in that we only assume query access to the platform (i.e., we don't have access to the platform's algorithm). At a high level, our framework creates different LLM agents that are seeded with original personalities from data from PEW research. Each of these agents is then allowed to interact with the contents of the social media platform according to their personality. We collect and analyze data about how the algorithm recommends content to different users and audit for a variety of different metrics.

3 - Formalizing Fairness With Respect To Stereotypes

Judy Hanwen Shen, Stanford University, Palo Alto, CA, United States

Machine learning algorithms used for high-stakes decision-making such as hiring must be carefully examined for unintentional biases. While existing group notions of fairness allow discrimination to be formally examined across groups, representational harms can also arise from outcomes determined by stereotypical features. Despite the plethora of empirical examinations of stereotypes and their harms, current literature offers no formal characterization of fairness with respect to stereotypes. In our work, we address this gap by introducing notions of fairness with respect to stereotypes that extend existing notions of group fairness. We examine how to achieve fairness with respect to stereotypical associations when there are multiple stereotypes under different fairness notions. We also present impossibility results characterizing when stereotype-based fairness notions are mutually exclusive.

SA44

Summit - 436

Recent Development in Simulation Analysis Methodology

Invited Session

Simulation Society

Chair: Linyun He, Georgia Institute of Technology, Atlanta, United States

1 - Digital Twin Validation with Multi-Epoch, Multi-Dimensional Data

Linyun He, Georgia Institute of Technology, Atlanta, GA, United States, Luke Rhodes-Leader, Eunhye Song

This paper studies validation of a simulation-based process digital twin (DT). We assume that at any point the DT is queried, the system state is recorded. Then, the DT simulator is initialized to match the system state and the simulations are run to predict the key performance indicators (KPIs) of the system at the end of the time epoch of interest. Our validation question is if the distribution of the simulated KPIs matches that of the system KPIs at every epoch. Typically, these KPIs are multi-dimensional random vectors and are non-identically distributed across epochs making it difficult to apply the existing validation methods. We devise a multiple hypothesis test that compares the marginal and joint distributions of the multi-dimensional KPI vectors, separately, by transforming the multi-epoch data to identically distributed observations. We empirically demonstrate that the test has good power when the system and the simulator sufficiently differ in distribution.

2 - Efficient Uncertainty Quantification and Reduction for Over-Parameterized Neural Networks

Haofeng Zhang, Columbia University, New York, NY, United States, Ziyu Huang, Henry Lam

Uncertainty quantification (UQ) is important for reliability assessment and enhancement of machine learning models. In deep learning, uncertainties arise not only from data, but also from the training procedure that injects substantial noises and biases. These hinder the attainment of statistical guarantees and, moreover, impose computational challenges on UQ due to the need for repeated network retraining. Building upon the recent neural tangent kernel theory, we create statistically guaranteed schemes to principally characterize, and remove, the uncertainty of over-parameterized neural networks with very low computation effort. In particular, our approach, based on what we call a procedural-noise-correcting (PNC) predictor, removes the procedural uncertainty by using only one auxiliary network that is trained on a suitably labeled dataset, instead of many retrained networks employed in deep ensembles. Moreover, by combining our PNC predictor with suitable light-computation resampling and simulation-based ideas, we build several approaches to construct asymptotically exact-coverage confidence intervals using as low as four trained networks without additional overheads.

3 - Effective Training and Optimization of Neural-Network-Based Simulation Metamodels

Peter Haas, University of Massachusetts Amherst, Amherst, MA, United States, Wang Cen

In prior work we introduced simulation metamodels based on graph and generative neural networks (NNs). Such metamodels can capture the effect of both input-parameter values and simulation-model structure on the simulation output, and allow fast approximation of a range of performance measures using a single metamodel. We describe a novel procedure for efficiently training a NN metamodel and also describe improvements to our original NN architecture that both improve accuracy and allow exact solution of mixed continuous-discrete optimization problems via an MILP solver.

4 - Machine Learning-Assisted Stochastic Kriging Metamodel for Offline Simulation Online Application

Haoting ZHANG, University of California, Berkeley, Berkeley, CA, United States, Jinghai He, Rhonda Righter, Zuo-Jun Shen, Zeyu Zheng

For many real-time decision problems in complex stochastic systems, a decision-maker observes the system status online in real time and needs to immediately evaluate the performance of alternative decision choices based on the observed system status. Simulation metamodels can effectively support these needs by constructing a fast-to-evaluate mapping from the system status and decision to system performance, using offline-generated simulation samples. However, when the system status involves high-dimensional information and a large number of simulated samples, classical simulation metamodeling approaches, such as stochastic kriging, may face challenges in terms of model specification, computational complexities, and computer storage demands. To address these challenges, we propose using machine learning models to assist stochastic kriging in building metamodels offline, that then can support online applications. The machine learning models can capture the potential nonlinear dependence of the stochastic kriging parameters on the high-dimensional system status. We analyze standard properties, such as mean squared errors and uncertainty quantification, for the proposed machine learning-assisted metamodel, and show its consistency and asymptotic validity. We demonstrate its comparative advantage of our approach through numerical experiments.

5 - Optimizing Discrete Event Simulations for Urban Transport: Subgraph Parallelization via the Kemeny constant in New York City's Citi Bike Network

Silvano Bernabel, The Graduate Center, City University of New York, New York, NY, United States, Felisa Vazquez-Abad

Simulating multiple vehicles in a realistically city-sized model is a challenging problem. While discrete-event or agent-based models are appropriate, execution times for long simulations can be prohibitive. Our research explores an intriguing alternative that admits parallelization of the discrete event simulations to increase overall efficiency. We use the simulation of Citi Bikes in New York city as a case study. Our approach divides the graph that describes the topology of the stations into subgraphs. Simulations on the subgraphs, executed in parallel, represent mobility within the corresponding region. Since the various subgraphs have interconnections, we must reconstruct the overall dynamics to approximate the original problem. Rather than dividing the space arbitrarily, we propose a novel approach based on the origin destination (OD) matrix. In the case of public bikes, these circulate around the city from station to station as different passengers ride them. We consider the normalized OD matrix as a Markov kernel that describes the movement of the vehicles and use the Kemeny constant to assess the subgroups that describe the stationary distribution as a singularly perturbed Markov chain. Reconstruction of the original system depends on the type of measurement chosen. We provide mathematical analysis and prove the validity of the approach via simulation. We highlight through a real-world application in transportation under various scenarios for system layout, capacity, and demand, enabling stakeholders to make quicker decisions for financial and social benefits.

SA45

Summit - 437

Healthcare Operations & Behavioral Influences

Invited Session

Health Applications Society

Chair: Jingyao Huang, University of Missouri-Kansas City, Overland Park, KS, United States

1 - Understanding Queue Abandonment: The Role of Wait Information and Distraction

Adam Arveson, University of Cincinnati, Cincinnati, OH, United States, Kanix Wang, Yinghao Zhang, Craig Froehle

Despite significant advancements in predictive analytics and service delivery enhancement, individuals face long wait times in Emergency Departments. The continued presence of queues underscores an ongoing need for researchers and practitioners to extend their focus beyond merely enhancing efficiency practices to improving the waiting experience itself. Inadequate management of customer waiting periods can significantly diminish satisfaction levels and increase abandonment, so developing effective strategies to mitigate the adverse effects of these experiences is imperative. Drawing on Maister's (1985) well-known principles, two predominant strategies have emerged: enhancing transparency through communication about the wait, and employing distraction techniques to divert attention from the wait.

While these commonly used approaches are well-established, there is something of a trade-off: more communication about the wait tends to draw customers' attention to their waiting, potentially increasing displeasure, while distraction necessarily lowers the firm's ability to communicate specifically about the wait. Many firms attempt to find a balance, using a bit of both tactics. But is this the best approach? This as-yet-unanswered question underscores a nuanced challenge in deciding how to best balance information transparency with distraction techniques within the context of wait management.

To investigate this phenomenon, we designed an online, between-subjects experiment where individuals engage in a waiting task. The study design incorporated six treatments, each varying in the degree of wait time information and the level of distraction. This experimental setup, which used a declining payment structure, enabled a detailed observation of participants' abandonment behaviors in response to each specific treatment scenario.

2 - Utilizing Specialists and Nurse Practitioners on Improving Accountable Care Organizations Efficiency: An Optimization Framework

Mayukh Majumdar, University of San Diego, San Diego, CA, United States, Anupam Agrawal, Arun Sen, Chelliah Sriskandarajah

Accountable Care Organizations (ACOs) are networks of primary care providers or PCP-led operating under shared savings or loss arrangements. They focus on delivering high-quality care while effectively managing costs. Over time, while some ACOs have demonstrated success in cost and quality performance, others have struggled. Additionally, the shortage of PCPs has been a long-standing concern for them and is expected to worsen. In response to this and the growing demand for healthcare access, ACOs leverage the expertise of specialists and nurse practitioners (NP) to ensure that patients receive comprehensive healthcare services. To better understand how ACOs can effectively utilize them, we study the care provided by specialists and NPs in single and multi-period settings using an optimization framework. Under the single-period setting, we simplify the formulation and propose an algorithm. Using a computational experiment, we find that (i) an increase in specialists and overreliance on NPs beyond an optimal threshold can compromise ACO efficiency, and (ii) under two-sided contracts, ACOs prioritize NPs as costs fluctuate. To reduce analytical complexity and provide an actionable solution, we introduce a simple algorithm for the multi-period setting to maximize efficiency in every period. Our study offers valuable insights for practitioners and policymakers to improve the performance of ACOs.

3 - A Machine Learning Model to Predict Heart Failure Readmission: Toward Optimal Feature Set

Sonia Jahangiri, Rochester Institute of Technology, Rochester, NY, United States, Masoud Abdollahi, Ehsan Rashedi, Nasibeh Azadeh-Fard

Hospital readmissions for heart failure patients remain high despite efforts to reduce them. In this study, we aimed to develop a machine learning based prediction model leveraging a nationwide hospitalization database to predict 30-day heart failure readmissions and to find the optimal feature set that leads to the highest AUC value in the prediction model. Heart failure patient data was extracted from the 2020 Nationwide Readmissions Database. A heuristic feature selection process incrementally incorporated predictors into logistic regression and random forest models, which yields a maximum increase in the AUC metric. Discrimination was evaluated through accuracy, sensitivity, specificity, and AUC. A total of 566,019 discharges with heart failure diagnosis were recognized. The feature selection method resulted in the identification of optimal feature sets including 20 and 22 variables from a pool of 30 and 31 features for the same-cause and all-cause datasets. Key predictors included age, payment method, chronic kidney disease, disposition status, number of comorbidities, and post-care encounters. This study provides meaningful insights into predictive modeling methodologies and influential features for forecasting heart failure readmissions.

SA46

Summit - 438

Models for Medical Decision Making and Health Policy

Invited Session

Health Applications Society

Chair: Sze-chuan Suen, University of Southern California, Los Angeles, CA, United States

Co-Chair: Yiwen Cao, University of Southern California, Los Angeles, CA, United States

Co-Chair: Citina Liang, University of Southern California, los angeles, CA, United States

1 - Optimizing Vehicle Routing and Nurse Scheduling for Mobile Vaccine Team Units in Los Angeles

Jing Jin, University of Southern California, Los Angeles, CA, United States, Suyanpeng Zhang, Sze-chuan Suen

Mobile Vaccination Team (MVT) units in Los Angeles County Department of Public Health (LACDPH) must simultaneously assign nurses to teams and route these teams to geographically dispersed homes to vaccinate patients. We formulate a mixed integer linear program (MILP) to minimize nurse travel time while constrained by appropriate nurse staffing requirements, nurse availability, patient scheduling preferences, and traffic congestion on LAC roads. Despite the inherent NP-hard nature of such scheduling and routing problems, we speed up the optimization process by using Big-M constraints and efficient formulation of constraints. We create code that takes inputs including the numbers of nurses, their home addresses, depot location, home visit addresses, and available time windows and durations for each home visit. It outputs a detailed scheduling sheet that specifies time windows, assigns nurse teams, and designates a team leader for each event. We parameterize the model with example data provided by the LACDPH and solve using Gurobi in Python. Our model efficiently solves scenarios ranging from small-scale (three nurses and five events) in seconds to larger real-world situations (50 nurses and over ten events) within 20 minutes. This indicates that typical weekly LACDPH scheduling could use this tool to find results in reasonable time.

2 - Non-Monotonicities in Variation with Transmission Rates in Risk-Stratified Compartmental Models: Implications for Infectious Disease Modeling

Jinghong Miao, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen

Compartmental models are widely employed to describe infectious disease transmission dynamics, and extreme value analysis with respect to transmission rates are often conducted as a follow-up step to predict the best- and worst- case scenario under model assumptions. We consider a stratified compartmental SIS (susceptible-infectious-susceptible) model with high-risk and low-risk subpopulations with the same clearance rate in both groups. We prove analytic properties and conduct sensitivity analyses given a known range for the number of contacts for high- and low-risk cohorts, which linearly influences intra- and inter-group transmission rates. We consider four metrics of epidemic severity – basic reproduction number R_0, the marginal increase in incidence per infected case (a proxy for the effective reproduction number R_t), endemic steady state prevalence, and cumulative prevalence. We establish that while R_0 and steady-state prevalence are monotonic with respect to intra- or inter-group transmission rates, R_t proxy and cumulative prevalence are non-monotonic with respect to transmission rates. It is thus not sufficient to only consider the minimum and maximum transmission rates for extreme value analyses for disease outcomes.

In this project, we highlight the counterintuitive fact that some reasonable metrics for epidemic severity are non-monotonic with respect to intra- or inter-group transmission rates in the two-group-risk-stratified SIS compartmental model. We believe this result can provide insights for epidemic control and cost-effectiveness analysis.

3 - Targeted Priority Mechanisms in Organ Transplantation

Ruochen Wang, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Burhaneddin Sandikci, Matthew Ellis

The persistent imbalance between organ supply and demand poses a significant challenge for the life-saving treatment of transplantation. This study introduces innovative targeted priority mechanisms inspired by the Eurotransplant Senior Program (ESP), designed to bridge the supply-demand gap while optimizing the matching between organs and recipients without mandating offer acceptance. Using a comprehensive queueing model, our study analyzes the strategic decisions of candidates under targeted priority mechanisms in equilibrium, establishing the optimal program design. Our evaluation of these mechanisms' impact on social welfare underscores their potential to improve overall welfare without adversely affecting any patient group. Furthermore, our simulation findings highlight the mechanisms' potential to increase yearly transplants by 1,000 and prevent up to 200 waitlist deaths annually.

4 - Predicting Eventual COVID-19 Vaccination Uptake by Vaccination Attitudes over TIME

Citina Liang, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen

Our study aims to predict COVID-19 vaccine uptake as of May 2023 by analyzing self-reported vaccination likelihood scores from April to December 2020. Utilizing the Understanding America Survey data, we assessed the predictive power of various machine learning methods, including Latent Transition Analysis (LTA), and compared these to a naive approach based on average likelihood scores. Our findings indicate varying performances among the models; notably, some models demonstrated high sensitivity, suggesting a greater ease in predicting individuals who will eventually opt for vaccination over those who will not. This study highlights the potential of advanced analytical techniques in forecasting public health behaviors, thereby informing targeted vaccine promotion strategies.

5 - Multi-Class Prediction of Missed Appointments in Primary Care Considering Health Impact

Yifang Yan, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen, Wen-Jan Tuan

Missed appointments in primary care not only waste healthcare resources but also disrupt the continuity of care in clinical workflows, adversely impacting patient health outcomes. We utilized machine learning models to predict three mutually exclusive appointment outcomes: completed visit, late cancellation, and no-show. The model demonstrated outstanding performance in predicting the probabilities of three-class outcomes. However, determining the risk thresholds at which these probabilities are sufficient to make practical classification decisions is non-trivial, especially when balancing the utility of correct and incorrect decisions for hospitals and patients. This challenge is further amplified with three-class prediction, which involve multiple risk thresholds and interactions between thresholds. To address this, we developed a multi-class prediction and decision model. We established support vector machine (SVM) model with hinge loss function modified to incorporate two threshold problems (TTP). TTP includes an "inconclusive zone" to focus on the individuals with sufficient evidence. We extended the hinge loss in SVM from two-segment piecewise linear function into three-segment to accommodate the "inconclusive zone." Loss function was also tailored to balance the waste in healthcare resources and the impact on health outcomes. We formulated a convex optimization problem and solved it in dual space. We applied our model to patients with type II diabetes. Health outcome impact was estimated using a microsimulation model by simulating the effects of missed appointments on HbA1c levels and subsequent complications and measured by loss in quality-adjusted life years. We compared our model with existing decision models by simulating the total cost during operations.

SA47

Summit - 439

Stochastic Modeling for Decision Making in the Cancer Care Continuum

Invited Session

Health Applications Society

Chair: Iakovos Toumazis, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Co-Chair: Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

1 - Designing Tertiary Cancer Services Under Burden Uncertainty

Abel Sapirstein, Georgia Institute of Technology, Atlanta, GA, United States, Lauren Steimle, Mathieu Dahan

Cancers are a growing cause of morbidity and mortality in low- and middle-income countries (LMICs). Geographic access plays a key role in both treatment and diagnosis - however a dearth of existing facilities often means that the demand for cancer services is not well understood. Using Rwanda as a case study, we adopt a distributionally robust optimization approach to simultaneously identify existing burdens and maximize access to treatment. First, we present a novel-path based formulation of two stage decision problem. In this problem's first stage a decision maker must simultaneously choose a set of facilities at which to add a set of tertiary cancer services and design a transportation network to improve patient access. In the second stage, patients use the transportation network to seek care at the place facilities. We then extend this formulation to instances in which demand is unknown and present preliminary work on solving large scale instances of this problem using Branch and Price and strengthening cuts.

2 - AI for Clinical Decision-Making in Oncology

Issam El Naqa, Moffitt Cancer center, Tampa, FL, United States

Artificial intelligence (AI) and Machine learning (ML) algorithms are revolutionary technologies that have captured the imagination of the public recently. They are currently transforming biomedical research, especially in the context of decision-making outcome modeling in cancer. These technologies have tremendous potentials for automating workflow, personalizing care, and reducing health disparity, to name a few prospects. Despite this excitement, only a few AI/ML models have been properly validated and fewer have become regulated products for routine clinical use. This can be attributed to multitude of concerning issues regarding the deployment of AI/ML driven technologies into the clinic. These concerns include technical, practical, ethical, and legal issues. In addition to skepticism related to commercialization hype, under representative data, inherent implementation bias, lack of robustness, and absence of prediction transparency. In this work, we will discuss some of these impending challenges, present general requirements for successful AI/ML implementation in the clinic, and highlight different approaches for detecting and mitigating explicit and implicit bias. We further show examples of applying these approaches in oncology applications from our work and others and discuss their implications to pave the way for safe and beneficial AI/ML decision making in clinical practice.

3 - Estimating sojourn time and sensitivity of screening for ovarian cancer using a Bayesian framework

Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Jiangong Niu, Martin Tammemagi, Ehsan Irajizad, Yu Shen, Karen Lu, Larissa Meyer, Iakovos Toumazis

Background: Ovarian cancer is the second leading cause of cancer-related deaths among all gynecologic cancers. Previous ovarian cancer screening trials failed to show significant mortality benefit. Estimating sojourn time, the time length from when the cancer becomes screen-detectable until clinical detection, is key to developing more effective future screening programs.

Methods: We modeled ovarian cancer progression as continuous-time Markov chain. The modality-specific sojourn time and screening's sensitivity were estimated from model parameters using a Bayesian approach to reproduce the outcome of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening and the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) trials. The quality of the estimates was validated using the posterior predictive p-value (PPPV). We also derived histology-specific sojourn times by adjusting the overall sojourn time based on their corresponding survival from SEER.

Results: The overall ovarian cancer sojourn time was estimated as 2.1 years (PPPV=0.469) in PLCO, and 2.0 years (PPPV=0.532) in the multimodal arm of UKCTOCS. The overall sensitivity of screening was 65.7% in the PLCO and 93.2% in the UKCTOCS. The stage-specific sensitivities of multimodal screening in PLCO were estimated as 39.1% and 82.9% for early and advanced stage disease, respectively. The histology-specific sojourn times ranged from 0.8-1.8 years for Type II and 2.9-6.6 years for Type I ovarian cancers.

Conclusions: Stochastic modeling and the Bayesian approach estimated reliable estimates of sojourn time and screening sensitivity. Annual screening is not effective in reducing ovarian cancer mortality because the sojourn time of the disease is too short.

4 - Personalized Rescheduling of Adaptive Organ-at-Risk-Sparing Radiation Therapy for Head and Neck Cancer under Re-planning Resource Constraints: A Novel Application of Markov Decision Processes

Fatemeh Nosrat, Rice University, Houston, TX, United States, Cem Dede, Lucas B. McCullum, Raul Garcia, Abdallah S. R. Mohamed, Jacob G. Scott, James E. Bates, Brigid A. McDonald, Kareem A. Wahid, Mohamed A. Naser, Renjie He, Amy C. Moreno, Lisanne V. van

Dijk, Kristy K. Brock, Jolien Heukelom, Seyedmohammadhossein Hosseinian, Mehdi Hemmati, Andrew Schaefer, Clifton D. Fuller

Background and Purpose: The purpose of this study was to determine personalized optimal timing for re-planning in adaptive organ-at-risk-sparing radiation therapy under limited re-planning resources in patients with head and neck cancer (HNC).

Materials and Methods: A novel Markov decision process (MDP) model was developed to determine optimal timing of re-plannings based on the patient's expected toxicity, characterized by normal tissue complication probability (NTCP), for four toxicities: xerostomia, dysphagia, parotid gland dysfunction, and feeding tube dependency at 6 months post-treatment. The MDP parameters were derived from a dataset comprising 52 HNC patients treated at the University of Texas MD Anderson Cancer Center between 2007 and 2013. Optimal re-planning strategies were obtained when the permissible number of re-plannings throughout the treatment was to 1, 2, and 3.

Results: The model recommended re-planning when the difference between the planned and actual NTCPs (ΔNTCP) was greater than or equal to 1%, 2%, 2%, and 4% at treatment fractions 10, 15, 20, and 25, respectively, exhibiting a temporally increasing pattern. The ΔNTCP thresholds remained constant across the number of re-planning allowances (1, 2, and 3). This result underscores the importance of re-planning for patients experiencing a slight change in ΔNTCP at fraction 10.

Conclusions: In contrast to prior work that relies on a single re-planning allowance or predetermined time intervals using a one-size-fits-all approach, the MDP model proposed in this paper offers a personalized, resource-aware, and scalable decision-making tool to identify optimal dynamic re-planning schedules, tailored to individual patients's needs.

SA48

Summit - 440

Advances in Sustainable Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Deniz Akturk, Washington University in St. Louis, 1 Snow Way Dr, St. Louis, MO, 63130, United States

Co-Chair: Zuguang Gao, University of California, Irvine, Irvine, CA, United States

1 - A Heuristic for near-Optimal Advertising Waterfalls

Jiamei Zhang, University of California, Irvine, Irvine, CA, United States, John Turner

Digital display advertising constitutes a significant source of revenue for online publishers. In cases when premium advertisers have priority over impressions, allocating ads using a waterfall strategy may be used. Through waterfall protocol, publishers sequentially solicit acceptance from advertisers at pre-set prices until the first acceptance. This study introduces a heuristic method to optimize the solicitation sequence under waterfall mechanisms. Given the exponential increase in potential strategies with more advertisers, optimizing these strategies is essential yet challenging for maximizing publisher revenue.

To address the need for rapid ad placement decisions upon content requests, we propose a stochastic optimization framework subject to complexities of uncertain response latencies and stringent loading time constraints for each impression. Our methodology incorporates dynamic programming and analytical models to structure waterfall strategy, and devise a heuristic scoring system for adaptive ranking of ad-slot requests.

By integrating several computationally efficient upper bounds, our approach is demonstrated to yield near-optimal revenue results. striking a balance between performance and computational efforts. This work provides a practical solution that can enhance publisher revenue under sequential decision environments.

2 - Using Influencers to Persuade

Zi (Elaine) Ling, Booth School of Business, University of Chicago, Chicago, IL, United States, Christopher Ryan, Mengzhenyu Zhang

The practice of cooperating with social media influencers has become a mutibillion-dollar industry. In this paper, we show how companies can use influencers with their own independent reputations as a vehicle to indirectly persuade target consumers (followers of the influencer) to make purchase. The followers act as Bayesian learners, gathering information from both the influencer and other consumers who have already made purchases before deciding to buy or not. This means the followers' decisions are interconnected and not made independently. Our study provides insights into how this indirect persuasion works and its implications for marketing strategies.

3 - Can Producers be Testers? A Peer-to-Peer Testing Framework for Milk Cooperatives

Liying Mu, University of Delaware, Newark, DE, United States, Sameer Mehta, Milind Dawande, Vijay Mookerjee

In the milk supply chain in developing countries, the production process and the testing process have traditionally been separate -- individual farmers produce milk and sell it to milk cooperatives, and the cooperative employs a dedicated entity to test the quality of milk. Since individual testing is costly, the cooperatives often do not conduct comprehensive testing of individual milk samples -- this creates incentives for farmers to free-ride. In this paper, we examine the viability of the idea of milk producers (i.e., individual farmers) also performing the role of testers. To this end, we develop and analyze a peer-to-peer testing framework. We exploit the fact that, as a community, farmers possess a significant amount of institutional knowledge about milk quality and the common ways in which milk can be adulterated. Further, potential concerns related to farmers' rudimentary testing instruments are addressed by multiple peers conducting independent tests on each milk sample. We show that peer-to-peer testing can increase farmers' income -- this incentivizes more farmers to join the cooperative and improve the milk quality, which in turn further increases their income, thus generating a ``beneficial virtuous cycle''.

4 - Microgrid Generation and Storage Investments: Scenarios of Price-Demand Independence and Dependence

Jiarui Wei, The University of Texas at Dallas, Richardson, TX, United States, Fariba Farajbakhsh Mamaghani, Metin Cakanyildirim

In today's energy transition, microgrids (MGs) emerge as vital components of the electricity landscape, offering profitable and reliable energy solutions, particularly for institutions like universities, airports, and hospitals. These localized energy systems, comprising local generators and consumers, act as alternative and supplementary suppliers alongside the grid, supported by short-term battery storage capabilities. However, determining optimal generation and storage capacities for MGs presents significant challenges due to uncertainties in demand, market prices, and their interdependence. This paper addresses these complexities by maximizing MG profit through closed-form formulas for optimized capacity levels under various price-demand scenarios. Notably, it reveals the impact of positive dependence between price and demand on MG profitability and investment decisions, highlighting the importance of incorporating interdependencies in sustainable energy planning.

5 - Optimal Inspection and Maintenance Planning for Utility Power Grids to Reduce Wildfire Risk

Abolfazl Taghavi, UCLA Anderson School of Management, Los Angeles, CA, United States, Ali Fattahi, Sriram Dasu, Reza Ahmadi

Wildfires have a significant impact on the environment, society, and infrastructures. During the last decades, the behavior of wildfires has changed significantly and become more severe. In addition, climate change and wildland-urban interface expansion made the wildfire problem even more challenging than it was in the past. As a result, wildfire management is becoming significantly important in the US and other countries. In 2018, PG&E, the largest utility company in California, almost went bankrupt because of igniting the campfire. Since then, utility companies have initiated several programs to reduce the fire ignition probability across their power grid, including maintenance and inspection of the power grid elements. In this project, we present an optimal inspection and maintenance policy for the utilities' power grid.

SA49

Summit - 441

Empirical Analysis of New Retail Models

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Antonio Moreno, Harvard Business School, Boston, MA, United States

1 - Hyper-Local Fulfillment Services: Consumer Demand and Service Design

Natalie Epstein, Harvard Business School, Cambridge, MA, United States, Santiago Gallino, Antonio Moreno

In this paper, we study the emerging industry of ultrafast grocery delivery services. Using data from a platform that provides ultrafast grocery delivery services in Latin America via a network of local micro-fulfillment centers, we find empirical evidence that customer demand is highly responsive to changes in delivery times (even when those times are short) and to the offline options available to the customer. Drawing from queuing theory and using a very granular dataset of ultra-fast grocery delivery operations, we study the tradeoffs associated with the design of the network of micro-fulfillment centers supporting ultrafast grocery delivery services. Our analysis provides a framework that retailers and platforms can use to evaluate their network design and how such decisions depend on the offline characteristics of the region.

2 - Expedited Service Promises: Evidence from a Food Ordering Platform

Simin Li, Tulane University, New Orleans, LA, United States, Nil Karacaoglu, Luyi Yang

We empirically investigate the intended and unintended consequences of expedited service promises on a food ordering platform. Employing a quasi-experimental design, we discover that the expedited service promise (ESP) leads to an increase in sales. However, it also alters the mix of order urgency, resulting in more ASAP orders and fewer advance-scheduled orders being placed. Indeed, utilizing change-in-changes analysis, we observe a leftward shift in the distribution of purchase to request delivery times. Consequently, we find an increase in the probability of customers receiving ASAP orders later than their requested time, as well as an increase in the probability of customers receiving advance-scheduled orders at least 15 minutes earlier than their requested time, resulting in food becoming cold.

This observation suggests that, in order to avoid delays for more urgent orders, servers may need to prioritize clearing out advanced orders early to accommodate urgent order processing. Nevertheless, the probability of late orders still increases. Furthermore, we empirically demonstrate that missed delivery time windows—both too late and too early—diminish customer retention on the platform. These findings carry significant managerial implications regarding how platforms should implement ESP to maximize profit increase. For example, leveraging admission controls or pricing delivery promises in conjunction with ESP could be effective strategies.

3 - The Value of Online Interactions for Store Execution

Borja Apaolaza, The Wharton School - University of Pennsylvania, Philadelphia, PA, United States, Felipe Caro, Victor Martínez-de-Albéniz

Problem definition: Omnichannel retailers interact with customers both online and offline. So far, they have used this richer information to optimize the sales process by designing the right channel and supply chain structures, and by personalizing offer, pricing and promotions. We advance an additional dimension of omnichannel value: retailers can use online clickstreams to better understand customer needs and optimize store layouts to maximize conversion. **Methodology/results:** We develop a model where in-store purchases depend on the customer's shopping list, and the effort required to locate and reach the products within the store. Category location in the store thus drives conversion. We then apply our model to a large home improvement retailer and find that shoppers' preferences are revealed by nearby online traffic, and hard-to-reach locations lead to lower conversion. Finally, we optimize category-location assignments using our demand model and find that putting higher-interest and higher-price items in the most effective locations can increase revenues by about 2% in comparison with models that ignore online clicks. **Managerial implications:** We show how using online clickstream information for optimizing offline operations can create significant value. More fundamentally, our results provide a word of caution that in some retail segments like home improvement, longer in-store paths might not necessarily be better.

4 - Disclosing Low Product Availability: An Online Platform&Rsquo;S Strategy for Mitigating Stockout Risk

Dmitry Mitrofanov, Boston College, Chestnut Hill, MA, United States, Benjamin Knight

Ensuring product availability and the successful fulfillment of orders are key priorities for any company operating in the retail industry. In this paper, we investigate how sharing information regarding the low availability of certain items can influence customers' purchase decisions, both preventing stockouts and mitigating the negative effects of stockouts with respect to customer long-term behavior. It is hard to predict the net impact of sharing item availability information on business metrics ex-ante because there are multiple effects that might act in opposite directions. The information-sharing policy could lead customers to not purchase items that are low in availability because those customers are averse to stockouts. Contrariwise, this approach might increase the popularity of low-availability items because consumers' cognitive bias leads them to place a higher value on items that are scarce.

In a field experiment, we exogenously share low item availability information with a random subset taken for a sample of more than 840K customers using Instacart. We find evidence that customers are 25% less likely to purchase low-availability items when item availability information is disclosed. In addition, we show that low product availability disclosure positively affects the platform's fundamentals in the context of our field experiment both in the short and long run. More specifically, our results indicate that this innovative and cost-free approach leads to a 5.33% increase in revenue per customer and a 4.9% increase in order frequency over the long term.

SA50

Summit - 442

Emerging-Market Retail Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Jiwen Ge, Dongbei University of Finance and Economics, Dalian, 116025

1 - Integrating Prediction and Optimization to Improve In-store Fulfillment Operations for Online Orders

Yang Zhan, Zhejiang University of Technology, Hangzhou, China, People's Republic of, Xueqi Wu, Zhi-Long Chen

This study addresses the critical challenge of optimizing order fulfillment in fast-paced online grocery markets. We develop an innovative integrated approach combining machine learning prediction (GBDT) with advanced optimization techniques to tackle heterogeneous picker efficiency and varying batch priorities. Our method includes a polynomial-time algorithm for a special case with a proven worst-case scenario and a heuristic with a theoretical worst-case bound for the general problem. Computational experiments using real-world data demonstrate significant improvements over current practices, reducing objective values by up to 23% and enhancing batch completion times. We provide crucial managerial insights on workforce strategies, emphasizing the value of experienced, full-time workers in boosting operational efficiency.

2 - Last-Mile Efficiency in Emerging Markets

Camilo Mora, Tecnologico de Monterrey, Monterrey, Mexico, Leopoldo Eduardo Cardenas-Barron, Jan Fransoo

Urbanization in emerging markets has led to market fragmentation, meaning a retail landscape dominated by nanostores that rely on distributors to serve them frequently to restock their businesses. In this context, a better understanding of last-mile operations is crucial to improving the efficiency of the distribution of goods. This article aims to advance research by providing data-driven evidence on the factors influencing dwell time in last-mile distribution for nanostores in emerging markets. We employed an observational research design in which 1000+ undergraduate students were trained to observe and document last-mile operations for nanostores without interfering with them, collecting over 4.5K observations. This research uses statistical analysis to shed light on the factors influencing dwell time and provides insights into freight vehicle composition in emerging markets. Our results show evidence that nanostores are served by a heterogeneous fleet, with over half being more than ten years old. We also capture that one third of the freight vehicles come from the informal sector. Regarding the operations, freight drivers were found to illegally park in 22% of the cases and that legal parking increases dwell time by 23%. These findings suggest a correlation between parking behavior and dwell time efficiency, emphasizing the impact of parking practices on logistics operations. We also found that companies can reduce 45% dwell time by prioritizing van sales over pre-sales strategies and using side doors to speed up unloading. The results hold implications for policymakers and companies seeking to elevate the efficacy of last-mile operations to serve nanostores.

3 - value of exclusive doorstep delivery in the last-100-meter distribution

Jiwen Ge, Dongbei University of Finance and Economics, Dalian, China, People's Republic of, Yang Zhan, Zheng Zhang

This paper investigates the value of exclusive doorstep delivery in the last-100-meter distribution service on e-commerce sales via a natural experiment conducted by Alibaba. By leveraging difference-in-differences models, we show significant sales change for Alibaba and its competitors. We also uncover the mechanism which drives the sales change.

SA51

Summit - 443

Emerging Topics in iFORM

Invited Session

MSOM: iForm

Chair: Yuqian Xu, UNC-Chapel Hill Kenan-Flagler Business School, Chapel Hill, NC, United States

Co-Chair: Youngsoo Kim, Loyola University Chicago, Tuscaloosa, AL, United States

1 - The Role of Late Payment Flexibility in Supply Chains

John Birge, University of Chicago, Chicago, IL, United States

Trade credit is an important source of financing in supply chains but the observed contractual terms are often not followed explicitly. Instead, terms of payment are often flexible and allow supply chain partners to share risk. This talk will review the role of this flexibility in supply chains and present empirical findings demonstrating its value.

2 - When the Dust Settles: Financing the Future of Land Reclamation

Qi Wu, Case Western Reserve University, Cleveland, OH, United States, Peter Ritchken

A key method to ensure that mining and oil companies restore land after use involves collecting taxes from operating firms, which are then placed in trust funds specifically designed to finance cleanup efforts once extraction activities end. This study examines regulatory frameworks designed to establish these funds. The sources of these contributions could be constant annual fees, payments tied to the volume of production, or royalty fees based on the revenue or profit generated by the operation. We analyze the four prevalent regulatory alternatives, all designed to be actuarially fair. The mining firms respond to each of these policies by adjusting their operating policies. Consequently, they are not indifferent between the policies. We characterize the properties of the firm's value and social welfare that lead to preferences over the four policies. We show that these preferences depend on firm's financial leverage and characteristics of production technology.

3 - Financing Supply Chain Resilience

Sriraman Sairam, Technical University of Munich, Heilbronn, Germany, David Wuttke, Andreas Kilian Gernert, Volodymyr Babich

We examine how financing arrangements and frictions affect supply chain resilience investments. In our game-theoretic model of a buyer, a supplier, and a bank, the supplier can invest in resilience, which reduces the supplier's production losses in case of a shock. The supplier is financially constrained and faces two financing frictions: moral hazard costs (the resilience investment is unobservable) and bankruptcy costs (future cash flows are lost in bankruptcy). We compare two financing arrangements: Under commercial loan financing, the supplier requests a loan from the bank; under buyer-intermediated financing (BIF), the supplier also obtains a loan from the bank, but the buyer guarantees the repayment and proposes the loan terms. Under commercial loan financing, we find that moral hazard costs can lead to credit rationing, limiting the supplier's resilience investments. In contrast, bankruptcy costs can accentuate these investments and change the direction of the moral hazard effect on investments. Bankruptcy costs can even motivate the supplier to invest more in resilience than she would without financing frictions. The buyer benefits from resilience and offers BIF only if it mitigates credit rationing. When BIF is offered, it always increases resilience. Surprisingly, financing frictions can increase the expected value for one of the firms: either the buyer or the supplier. The supplier's bankruptcy costs benefit the buyer when the supplier uses a commercial loan. Moral hazard costs can benefit the supplier when they motivate the buyer to offer BIF.

4 - Role of Liquidity in Retail Settings

Panos Kouvelis, Washington University in St. Louis, St.Louis, MO, United States, Naveed Chehrazi, Wenhui Zhao

We present an interesting model capturing consumer liquidity concerns on retail pricing and inventory decisions.

SA52

Summit - 444

Reinforcement Learning and Causal Inference with Applications in Service Systems

Invited Session

MSOM: Service Operations

Chair: Yao Xie, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: He Wang, Georgia Tech, Atlanta, GA, United States

1 - Reinforcement Learning in Latent Heterogeneous Environments

Elynn Chen, New York University, New York City, NY, United States, Rui Song, Michael Jordan

Reinforcement Learning holds great promise for data-driven decision-making in various social contexts, including healthcare, education, and business. However, classical methods that focus on the mean of the total return may yield misleading results when dealing with heterogeneous populations typically found in large-scale datasets. To address this issue, we introduce the K-Value Heterogeneous Markov Decision Process, a framework designed to handle sequential decision problems with latent population heterogeneity. Within this framework, we propose auto-clustered policy evaluation for estimating the value of a given policy and auto-clustered policy iteration for estimating the optimal policy within a parametric policy class. Our auto-clustered algorithms can automatically identify homogeneous subpopulations while simultaneously estimating the action value function and the optimal policy for each subgroup. We establish convergence rates and construct confidence intervals for the estimators. Simulation results support our theoretical findings, and an empirical study conducted on a real medical dataset confirms the presence of value heterogeneity and validates the advantages of our novel approach.

2 - Different Faces of Difference-in-Differences

Anqi Zhao, Duke University, Durham, NC, United States

In many social science applications, researchers use the difference-in-differences (DID) estimator to establish causal relationships, exploiting cross-sectional variation in a baseline factor and temporal variation in exposure to an event that presumably may affect all units. This approach, often referred to as generalized DID (GDID), differs from canonical DID in that it lacks a "clean control group" unexposed to the event after the event occurs. In this paper, we clarify GDID as a research design in terms of its data structure, feasible estimands, and identifying assumptions that allow the DID estimator to recover these estimands. We frame GDID as a factorial design with two factors: the baseline factor, denoted by G, and the exposure level to the event, denoted by Z, and define effect modification and causal interaction as the associative and causal effects of G on the effect of Z, respectively. We show that under the canonical no anticipation and parallel trends assumptions, the DID estimator identifies only the effect modification of G in GDID, and propose an additional generalized parallel trends assumption to identify causal interaction. Moreover, we show that the canonical DID research design can be framed as a special case of the GDID research design with an additional exclusion restriction assumption, thereby reconciling the two approaches. We illustrate these findings with empirical examples from economics and political science, and provide recommendations for improving practice and interpretation under GDID.

3 - Multi-Agent Reinforcement Learning for Joint Police Patrol and Dispatch

Matthew Repasky, Georgia Institute of Technology, Atlanta, GA, United States, He Wang, Yao Xie

Modern service systems are often composed of patrol and dispatch phases. Many methods for optimizing service system operations focus on optimizing the patrol or dispatch actions alone, neglecting to consider a joint decision-making process. In this work, we develop a joint optimization scheme for patrol and dispatch decision makers. Our method targets a flexible objective using deep multi-agent reinforcement learning (RL) and mixed-integer programming, formulating the patrol and dispatch problem in terms of patrollers and incidents on a graph. Our work builds upon prior RL approaches focused on selecting within combinatorial action spaces, extending such approaches to learn a dispatcher which pairs patrollers to incidents. Simultaneously, Q-learning with parameter sharing is conducted to optimize cooperative patrol agents. An alternating procedure is developed to jointly optimize the parameters of the dispatcher and patrollers.

We demonstrate that this approach results in more rapid response than optimizing the patrol or dispatch policies alone when optimizing an efficiency-based objective. Furthermore, the flexible reward definition also permits targeting alternative objectives, such as those encouraging equity. We show that optimizing with respect to equity may yield policies which are more fair than those optimized naively for efficiency, resulting in less disparity in response time to disproportionate groups. Finally, our method is applied to a realistic setting based upon three Atlanta Police Department (APD) beats in Southwest Atlanta. We further exhibit the benefit of our method by comparing it to a realistic policy based upon GPS data of APD patrollers.

4 - Counterfactual Generative Models for TIME-Varying Treatments

Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

Estimating the counterfactual outcome of treatment is essential for decision-making in public health and clinical science, among others. Often, treatments are administered in a sequential, time-varying manner, leading to an exponentially increased number of possible counterfactual outcomes. Furthermore, in modern applications, the outcomes are high-dimensional and conventional average treatment effect estimation fails to capture disparities in individuals. To tackle these challenges, we propose a novel conditional generative framework capable of producing counterfactual samples under time-varying treatment, without the need for explicit density estimation. Our method carefully addresses the distribution mismatch between the observed and counterfactual distributions via a loss function based on inverse probability re-weighting, and supports integration with state-of-the-art conditional generative models such as the guided diffusion and conditional variational autoencoder. We present a thorough evaluation of our method using both synthetic and real-world data. Our results demonstrate that our method is capable of generating high-quality counterfactual samples and outperforms the state-of-the-art baselines.

5 - Optimal Policy Learning in Nonstationary Markov Decision Processe

Liyan Xie, CUHK-SZ, Shenzhen, China, People's Republic of, Jie Meng, Minshuo Chen, Mengdi Wang, Vincent Poor

We consider a piecewise stationary Markov Decision Process in which different pieces are segmented by unknown change-points. We propose the Detection-UCBVI Algorithm, which introduces online change detection methods into an Upper Confidence Bound Value Iteration (UCBVI) type algorithm. The Detection-UCBVI algorithm can effectively identify change-points and thus allows the learning algorithm to restart more quickly after changes. We analyze regret in the finite-time episodic setting with K episodes for arbitrary sizes of changes. Numerical results are provided to demonstrate the advantage of Detection-UCBVI as compared with the vanilla UCBVI algorithm without detection.

SA53

Summit - 445

Emerging Topics on Digital Economy and Platforms

Invited Session

MSOM: Supply Chain

Chair: Xin Fang, Singapore Management University, Singapore, Singapore

1 - Multi-Product Pricing and Multi-Channel Budget Allocation for E-Commerce Retailers: a Data-Driven Target-Oriented Robust Optimization Framework

Qiyuan Deng, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Yue Zeng

As multi-channel marketing in online platforms becomes increasingly popular, e-commerce retailers face challenges in effectively distributing their budget across multiple marketing channels and setting prices for multiple products to maximize their total return. To tackle these challenges, we propose a data-driven decision-making framework using robust optimization to improve an e-commerce retailer's profitability through optimized multi-product pricing and multi-channel budget allocation. Based on data from a world-renowned retailer selling skincare products, we empirically construct a demand prediction model capturing two essential features of multi-channel marketing in online platforms – spillover effect and promotion effect. By incorporating the demand model, we develop a data-driven framework to optimize the e-commerce retailer's pricing and budget allocation decisions. Facing demand uncertainty, we adopt a target-oriented robust optimization (TRO) approach to absorb as much uncertainty as possible in our solutions while ensuring a target total return. Based on the same data set, the TRO approach substantially outperforms a deterministic policy, yielding an approximately 24% increase in the retailer's daily average return. The TRO approach is especially promising, yielding approximately 31% higher daily average return, during promotion phases when it is more difficult to predict demand. By analyzing the retailer's optimal pricing and budget allocation decisions, we provide practical guidance to e-commerce retailers. Specifically, retailers should lower product prices to increase their competitiveness in online platforms. Besides, during promotion phases, we recommend to allocate more budget to marketing channels that can increase product exposure to potential consumers, instead of marketing channels functioning as search ads.

2 - Managing the Personalized Order-Holding Problem in Online Retailing

Yun Fong Lim, Singapore Management University, Singapore, Singapore, Shouchang Chen, Zhenzhen Yan

A significant percentage of online consumers place consecutive orders within a short duration. To reduce the total order arrangement cost, an online retailer may consolidate consecutive orders from the same consumer. We investigate how long the retailer should hold the consumer's orders before sending them to a third-party logistics provider (3PL) for processing. In this order-holding problem, we optimize the holding time to balance the total order arrangement cost and the potential delay in delivery. We show that the optimal order-holding decisions follow a threshold-type policy: Hold any pending orders if the holding time is within a threshold, or send them to the 3PL otherwise. Whenever the consumer places a new order, the holding time is reset and the threshold is updated based on a cumulative set of her past consecutive orders in her shopping journey. Using a consumer's sequential decision model, we personalize the threshold by finding its closed-form expression in the consumer's order features. We determine the model's coefficients and evaluate the threshold-type policy using the data of the 2020 MSOM Data Driven Research Challenge. Extensive numerical experiments suggest that the personalized threshold-type policy outperforms two commonly-used benchmarks by having fewer order arrangements or shorter holding times. Furthermore, personalizing the order-holding decisions is significantly more valuable for "enterprise" customers. The threshold is higher for "plus" consumers, female consumers, and consumers in the age group of 16-25. The threshold for tier-1 cities is lower than that for tier-2 to tier-4 cities but higher than that for tier-5 cities.

3 - Optimal Growth of a Two-Sided Platform with Heterogeneous Agents

Yixin ZHU, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Hongfan Chen, Renyu Zhang, Sean Zhou

We consider the dynamics of a two-sided platform, where the agent population on both sides experiences growth over time with heterogeneous retention and new adoption. The compatibility between buyers and sellers is captured by a bipartite network. The platform strategically sets commissions to optimize its total profit over a horizon of T periods, considering the trade-off between short-term profit and growth as well as the spatial imbalances in supply and demand. We design an asymptotically optimal commission policy with the profit loss upper-bounded by a constant independent of T. To obtain this policy, we first develop a single-period benchmark problem that captures the optimal steady state of the platform, then delicately control the growth of the agent types with the lowest relative population level compared with the benchmark in each period. We further examine the impact of the growth potential and the market network structure on agents' payment/income, and the platform's optimal commissions and profit. We first show that for each agent type, higher ratios of their compatible counterparts' long-term growth potential to their own cause lower payments (higher income) at the optimal steady state. In addition, the impact of the relative long-run growth potential on the platform's optimal commissions in a submarket depends on the convexity/concavity of the value distribution function of agents. Finally, we show that a ``balanced'' network, where the relative long-run growth potential between sellers and buyers for all submarkets is the same as that for the entire market, allows the platform to achieve maximum profitability.

4 - To Keep Prices Consistent OR Not: Multi-Channel Retailing with Consumers' Fairness Concerns

Guang Xiao, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Xiaomeng GUO, Yumeng Li, Wenxin Xu

In this talk, we study a stylized game-theoretic model to investigate a multi-channel retailer's optimal pricing strategy in the presence of consumers' fairness concerns. We investigate the impact of such concerns on the firm's pricing strategy, profitability, and consumer surplus. Our results can provide useful managerial insights for multi-channel retailers to adopt proper pricing strategies when facing fairness-concerned consumers (This research is supported by the Research Grants Council of Hong Kong [General Research Fund Grant: PolyU 15505621]).

5 - Learning and Forgetting in Gig Work: Behavior-Based Task Allocation in An on-Demand Grocery Platform

Reeju Guha, CUNEF Universidad, Madrid, Spain, Daniel Corsten, Maqbool Dada

Online platforms operating on a gig-contractor model rely on self-scheduling gig workers to meet real-time demand, while maintaining desired performance levels. Most platforms allocate tasks randomly based on proximity of the worker to the store, or based on prior ratings. However, such allocations are often not useful as gig workers frequently have gaps in continuity of their service, thereby leading to a 'forgetting' effect, especially when they visit unfamiliar stores after a long service-gap. Furthermore, allocating tasks on the basis of prior ratings could cause an 'experience-mismatch' where higher-ranked workers gather more experience than their 'inexperienced' counterparts. Therefore, we investigate whether recent (intra-day) experience could mitigate the detrimental effects of 'forgetting' on workers' performance, and therefore be utilized as a ranking parameter in task allocations. Utilizing data from an online grocery platform, we develop an econometric model to analyze workers' performance, based on their intra-day and prior task experience, while accounting for selection bias and endogeneity. We further examine the moderating effects of task batching, complexity, and discretion. Our findings reveal that as intra-day and prior experience increase, both productivity and service quality improves. We further observe that excessive batching can improve productivity, but increase delays. Lastly, when complexity or discretion increases, productivity improves, but service quality reduces if intra-day experience crosses a threshold-limit. Utilizing our results, we rank gig-workers and develop a task-allocation algorithm to allocate work based on workers' rank, and task complexity, while adjusting for compensation. Finally, we calculate predicted improvements in performance from the new allocations.

SA54

Summit - 446

Frontiers in Sequential Learning and Experimentation

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Tu Ni, Harvard Business School, Boston, United States

1 - Online Uniform Risk TIMES Sampling: First Approximation Algorithms, Learning Augmentation with Full Confidence Interval Integration

Xueqing Liu, Duke-NUS Medical School, Singapore, Singapore, Kyra Gan, Esmaeil Keyvanshokooh, Susan Murphy

In digital health, the strategy of allocating a limited treatment budget uniformly across risk times is crucial to reduce user fatigue. This strategy, however, encounters a significant obstacle due to the unknown actual number of risk times, a factor not adequately addressed by existing methods lacking theoretical guarantees. This paper introduces, for the first time, the online uniform risk times sampling problem within the approximation algorithm framework. We propose a randomized approximation algorithm that enjoys robust theoretical guarantees across all problem instances or sample paths. Next, we extend our algorithm to the setting where a confidence interval [L,U] for the true risk time is provided, potentially through a machine learning algorithm. We conduct the competitive ratio analysis for our proposed learning-augmented approximation algorithm, demonstrating its consistency in the strong sense - optimal performance is achieved when the confidence interval width is zero - and robustness - the learning-augmented algorithm performs no worse than the non-learning augmented counterpart. We assess the performance of our proposed algorithms through numerical experiments conducted on both synthetic and real-world datasets. Without learning augmentation, our randomized algorithm outperforms the benchmark b/T. With learning augmentation, our learning-augmented algorithm performs no worse than both the naive benchmark b/U and the randomized algorithm.

2 - Adaptively Learning to Select-Rank in Online Platforms

Sharon Wang, Stern School of Business, New York University, New York, NY, United States, Perry Dong, Ying Jin, Ruohan Zhan, Zhengyuan Zhou

Ranking algorithms are fundamental to various online platforms across e-commerce sites to content streaming services. Our research addresses the challenge of adaptively ranking items from a candidate pool for heterogeneous users, a key component in personalizing user experience. We develop a user response model that considers diverse user preferences and the varying effects of item positions, aiming to optimize overall user satisfaction with the ranked list. We frame this problem within a contextual bandits framework, with each ranked list as an action. Our approach incorporates an upper confidence bound to adjust predicted user satisfaction scores and selects the ranking action that maximizes these adjusted scores, efficiently solved via maximum weight imperfect matching. We demonstrate that our algorithm achieves a cumulative regret bound of $O(d\sqrt{NKT})$ for ranking K out of N items in a d-dimensional context space over T rounds, under the assumption that user responses follow a generalized linear model. This regret alleviates dependence on the ambient action space, whose cardinality grows exponentially with N and K (thus rendering direct application of existing adaptive learning algorithms -- such as UCB or Thompson sampling -- infeasible). Experiments conducted on both simulated and real-world datasets demonstrate our algorithm outperforms the baseline.

3 - Optimal Statistical Inference in Adaptive Experiment Design: Private, Robust and Efficient

Jiachun Li, Massachusetts Institute of Technology, Cambridge, MA, United States, David Simchi-Levi

It's widely known that using adaptive experimental design can use the historical data to enhance data efficiency and improve statistical power. However, the adaptive nature of data collection process makes the traditional statistical analysis fails. In this work, we try to understand the question of what's the statistical limit of any experimental design and whether we can design an adaptive experiment which attains this statistical lower bound. Besides, we investigate the privacy and robustness issue in adaptive experimental design and analyze their impact on the statistical power.

4 - Design-Based Confidence Sequences: A General Approach to Risk Mitigation in Online Experimentation

Dae Woong Ham, Michigan Ross Business School, Ann Arbor, MI, United States, Iavor Bojinov, Martin Tingley, Michael Lindon

Randomized experiments have become the standard method for companies to evaluate the performance of new products or services. In addition to augmenting managers' decision-making, experimentation mitigates risk by limiting the proportion of customers exposed to innovation. Since many experiments are on customers arriving sequentially, a potential solution is to allow managers to "peek" at the results when new data becomes available and stop the test if the results are statistically significant. Unfortunately, peeking invalidates the statistical guarantees for standard statistical analysis and leads to uncontrolled type-1 error. Our paper provides valid design-based confidence sequences, sequences of confidence intervals with uniform type-1 error guarantees over time for various sequential experiments in an assumption-light manner. In particular, we focus on finite-sample estimands defined on the study participants as a direct measure of the incurred risks by companies. Our proposed confidence sequences are valid for a large class of experiments, including multi-arm bandits, time series, and panel experiments. We further provide a variance reduction technique incorporating modeling assumptions and covariates. Finally, we demonstrate the effectiveness of our proposed approach through a simulation study and three real-world applications from Netflix. Our results show that by using our confidence sequence, harmful experiments could be stopped after only observing a handful of units; for instance, an experiment that Netflix ran on its sign-up page on 30,000 potential customers would have been stopped by our method on the first day before 100 observations.

5 - Geometry-Aware Approaches for Balancing Performance and Theoretical Guarantees in Linear Bandits

Yuwei Luo, Stanford University, Stanford, CA, United States

Recent research in the d-dimensional stochastic linear bandit literature highlights a discrepancy: algorithms like Thompson sampling and Greedy perform well empirically but have pessimistic theoretical regret bounds. This discrepancy arises because these algorithms may perform poorly in certain instances but excel in typical ones. To address this, we propose a data-driven technique that tracks the geometric properties of the uncertainty ellipsoid around the main problem parameter. This enables us to formulate an instance-dependent frequentist regret bound, incorporating geometric information, for a broad class of algorithms, including Greedy, OFUL, and Thompson sampling. Our approach identifies and corrects poorly performing instances, achieving minimax optimal regret of order $d\sqrt{T}$ for a T-period decision-making scenario, up to logarithmic factors, while maintaining the algorithms' empirical efficacy. We validate our findings with simulation results using synthetic and real data.

SA55

Summit - 447

Humanitarian and Disaster Operations Management

Invited Session

Public Sector OR

Chair: Christopher Zobel, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Andrew Arnette, Virginia Tech, Blacksburg, VA, United States

1 - Vulnerability-Based Prioritization in Disaster Response Efforts

Feyza Guliz Sahinyazan, Simon Fraser University, Vancouver, BC, Canada, A.Irfan Mahmutogullari, Halenur Sahin

Humanitarian logistics literature commonly uses Equity, Efficiency and Effectiveness (3E) objectives. The equity objective strives to minimize differences among individual treatments by assuming that everyone is equally affected by a disaster. Efficiency measures aim to reduce the costs of aid programs, while effectiveness focuses on the quality of humanitarian aid, measured by factors such as response time, reduced risk, or human suffering. The inherent assumption of 3E objectives is the homogeneity of the beneficiaries. However, it is essential to acknowledge that disasters disproportionately affect individuals from socioeconomically disadvantaged backgrounds. Vulnerable groups, including the low-income, elderly, disabled, or marginalized, typically encounter unique challenges and heightened risks during disasters. Any measure assuming homogeneous demand will overlook the intersectionalities experienced by vulnerable communities. This paper introduces an alternative measure prioritizing vulnerable populations in disaster response, aiming for a more inclusive and compassionate disaster management strategy. To compare the performance of this approach against the traditional 3E measures, we study the emergency assembly point allocation problem and use Istanbul's neighborhood-level vulnerability and population characteristics for our computational analyses. Our results demonstrate that the vulnerability-based prioritization approach can single-handedly achieve more inclusive results for vulnerable populations without significantly deteriorating 3E objectives and non-vulnerable population outcomes.

2 - An Information-Theoretic Framework for Tracking Changes in 311 Calls Using Transformer-Based Hierarchical Topic Modeling

Onur Seref, Virginia Tech, Blacksburg, VA, United States, Christopher Zobel, Michelle Seref

311 services provide a channel to connect municipal agencies with the public. 311 calls are typically recorded and organized under pre-defined categories. However, during events such as crises, natural language usage among affected individuals change significantly, and may render the existing categories insufficient due to the unforeseen changes in context. Transformer-based hierarchical topic models capture context through deep semantic relationships and organize them into hierarchical clusters as topics. These topics may naturally change over time while the categories remain static, which causes a disconnect. To study this problem, we introduce an information-theoretic framework to establish a relationship between the topics and the categories. We study 311 calls from Orlando, FL and the surrounding municipalities within a six month time period that spans three months before and three months after the onset of the COVID-19 outbreak. We track changes under each category and each topic to match them based on their mutual information. Using the hierarchical structure of the topics, we identify expansions, reductions, or shifts in context under affected categories, and suggest new alternatives for organizing categories.

3 - Efficiency Criteria for National Public Private Emergency Collaborations: Findings from the German Food Emergency Preparedness System

Marcus Wiens, Technical University Freiberg, Freiberg, Germany

The secure supply of essential goods to the population in the event of a crisis has been of great relevance even before Covid-19 and the war in Ukraine. The article provides information on the experiences and findings from the four-year NOLAN research project (as well as follow-up projects and ongoing stakeholder exchanges), which identified opportunities to make emergency food provision in Germany more robust and reliable with the help of a public-private emergency collaboration. It is shown that both companies and authorities place high value on efficiency in addition to the robustness and effectiveness of structures and processes.

On the basis of empirical surveys, network optimization and game-theoretical analyses, two focal domains are analyzed from the point of view of efficient structures and processes of emergency supply. The first focus is on the optimal design of emergency storage and examines the costs of different levels of integration of public emergency storage into the commercial system. The second focus examines the incentives of private companies to integrate relief processes required in an emergency into the commercial retail system. These include an early warning system, strategic co-location of warehouses, rush orders and the implementation of emergency deliveries through re-routing as well as the voluntary provision of a minimum emergency stock capacity.

4 - To Donate OR Not to Donate: Optimal Policies for Supermarket In-Kind Donations to Food Banks

Shabnam Salehi, The University of Alabama, Tuscaloosa, AL, United States, Irem Orgut, Emmett Lodree

Food waste is a significant problem with economic, environmental, and social impacts. Supermarkets view the donation of perishable items close to their expiration date to food banks as a mean to reduce waste-related costs and improve their social image. Food banks, as non-profit organizations, depend greatly on receiving these perishable donations promptly to manage their operations effectively and distribute food efficiently to those in need. However, supermarkets, with the goal of maximizing sales and minimizing waste, often donate perishable goods as close to their expiration date as possible, even though they receive tax credits for these donations. Our study focuses on the food bank supply chain from the standpoint of supermarkets, with the goal of maximizing their profit by developing a multi-period stochastic dynamic programming model that identifies the optimal timing and quantity for donating perishables under stochastic demand. The analysis reveals that delaying donations until the last minute is not always optimal for supermarkets. Under certain conditions, early donation strategies can lead to better outcomes, enhancing the supermarkets' profitability while simultaneously reducing food waste in the food bank supply chain more effectively. The paper emphasizes the importance of aligning the donation strategies of supermarkets with the objectives of food banks to maximize the efficiency and societal benefits of the hunger relief supply chain.

5 - A Multi-Commodity Approach to Long-Term Climate-Induced Displacement

Deniz Emre, University of Oklahoma, Norman, OK, United States, Kash Barker, Andrés David González

The visible effects of climate change are increasing and are expected to escalate over time. One significant consequence is the displacement of populations due to climate-induced events, a process that has already started and is projected to lead to mass movement in the foreseeable future. With this motivation, we focus on long-term relocations caused by climate-induced events such as drought and sea level rise. We develop a multi-commodity model representing distinct groups of individuals as separate commodities. By distinguishing different groups as "commodities", we analyze displacement patterns and provide tailored long-term planning strategies.

SA56

Summit - 448

Convergence of Multi-sectoral Operations Research

Panel Session

Women in OR/MS (WORMS)

Co-Chair: Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

Co-Chair: Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

1 - Panelist

Julie Ivy, University of Michigan, Ann Arbor, MI, United States

2 - Panelist

Maria Mayorga, North Carolina State University, Raleigh, NC, United States

3 - Panelist

Jessica Boakye, University of Massachusetts Amherst, Amherst, MA, United States

4 - Panelist

Renata Konrad, Worcester Polytechnic Institute, Worcester, MA, United States

5 - Panelist

Anahita Khojandi, University of Tennessee, Knoxville, TN, United States

6 - Panelist

Evin Jacobson, CDC, Atlanta, GA, United States

SA57

Summit - Terrace Suite 1

Revolutionizing Patient Care: The Fusion of Machine Learning and Optimization Techniques

Invited Session

Health Applications Society

Chair: Gian-Gabriel Garcia, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Himadri Pandey, Georgia Institute of Technology, Atlanta, 30308, United States

1 - A Monotonic Policy Q-Learning Algorithm for Hypertension Management

Mina Yi, Dartmouth College, West Lebanon, NH, United States, Wesley Marrero

Reinforcement learning algorithms offer a promising approach for addressing treatment planning in healthcare applications. Within the context of hypertension treatment, traditional algorithms often fail to generate policies that align with clinical intuition. Clinicians commonly favor treatment plans that adjust medication intensity over time in response to the patient's health status, gradually increasing it as needed. To address this requirement, we introduce a monotonic Q-learning algorithm to provide online, treatment policies that adapt to individual patients. This algorithm is evaluated over a set of patients and demonstrates comparable efficacy to traditional offline algorithms using mixed-integer programming.

2 - Robust AI-Assisted Radiotherapy Planning

Arkajyoti Roy, The University of Texas at San Antonio, San Antonio, TX, United States, Justin Boutilier, Ruiqi Li

Artificial intelligence (AI) based auto-contouring offers a quick and cost-effective solution for delineating structures on CT scans, and has the potential to increase standardization and accessibility of radiation therapy by lowering the time costs of the planning process. However, AI-based auto-contouring can produce errors, especially when the training data quality and/or quantity is limited. This can degrade the precision of each contour and reduce the effectiveness of the resulting treatment plan. Our solution to this problem is a predictive-prescriptive framework that fuses the AI-based predictions and its errors with a subsequent robust treatment planning model to produce plans that can withstand auto-contouring errors. Our results show sizable plan improvements in using the proposed framework, compared to deterministic approaches.

3 - Learning Optimal Prescriptive Trees for Substance use Treatment: Trade-Offs Between Fairness, Interpretability, and Accuracy

Han Kyul Kim, University of Southern California, Los Angeles, CA, United States, Phebe Vayanos

Substance abuse poses a significant societal challenge in the United States, particularly among adolescents and young adults. Despite the urgency, most treatment facilities assign treatments based on resource availability, often overlooking individual needs or backgrounds of patients.

In this work, we explore the impact of data-driven personalized treatment assignments using prescriptive trees. Based on the assumptions of causal inference, we train an optimal prescriptive tree from the GAIN dataset, the largest longitudinal dataset capturing real treatment outcomes. This approach allows for a more personalized treatment assignment policy that considers patient-specific needs while maintaining model interpretability.

Furthermore, we introduce the concept of the “price of interpretability", which quantifies the difference in expected treatment outcomes between black-box models and interpretable models, highlighting the trade-off between model performance and interpretability. Our work reveals disparities in the price of interpretability among gender groups in our dataset, indicating that some groups may face a higher cost in expected outcomes when using interpretable models. To address this disparity, we propose a modified prescriptive tree that aims to distribute the price of interpretability more equally among protected groups, balancing fairness with model interpretability in treatment assignments.

Our research seeks to empower clinicians and social workers by integrating data-driven decision-making into fair and interpretable treatment assignment policies. The ultimate goal is to enhance the effectiveness of treatment for adolescents and young adults battling substance abuse, fostering a more fair and personalized approach to care.

4 - Advancing Patient Monitoring Strategies in Tuberculosis Treatment Using Reinforcement Learning

Maryam Kheirandish, University of Arkansas, Fayetteville, AR, United States, Shengfan Zhang

Health systems are currently facing unprecedented financial challenges. Both the overuse and underuse of tests deplete essential health resources and impede timely diagnosis and treatment. A primary driver of test overuse is the inherent uncertainty in medical decision-making. Merely ordering additional diagnostic tests does not necessarily prevent physicians from missing a condition. To counteract test overuse, a predictive model can assist physicians by providing insights into a patient's current health status, based on existing test results and associated uncertainties. This enables physicians to make informed decisions about whether additional tests are necessary or if treatment can proceed based on the existing data, ensuring conditions are not overlooked. Test overuse occurs in both disease diagnosis and treatment processes. While predictive uncertainty may be sufficient for decision-making at the diagnosis stage, a sequential decision-making approach is crucial for ongoing treatment monitoring. In this study, we introduce a contextual reinforcement learning framework that integrates uncertainty quantification in outcome prediction models to improve treatment monitoring strategies. This framework aims to identify scenarios where additional monitoring tests can enhance the predictive uncertainty of treatment outcomes. The model considers the impact of test ordering on current predictive uncertainty and on subsequent predictions used in future patient follow-ups. We apply this model to enhance the treatment monitoring of tuberculosis patients.

5 - Resource Allocation

Himadri Pandey, Georgia Institute of Technology, Atlanta, GA, United States

Concussions, the most prevalent form of traumatic brain injuries, pose a significant public health concern, contributing to 1.6-4 million sports and recreation-related incidents annually. Alarmingly, half of these incidents remain unreported or undetected. Swift and precise concussion diagnosis is imperative for effective injury management, as it reduces the potential for severe short-term and long-term consequences like cognitive impairment, depression, and neurodegenerative disorders. Baseline testing has emerged as a promising tool in diagnosing concussions. Although it is particularly beneficial for certain subgroups, there is a scarcity of such resources. Often, they are allocated to high-impact sports like football, disregarding other demographics that may have a greater need. Baseline testing is especially useful for specific populations compared to standard reference points. This presents a dilemma: given the limited availability of baseline testing, which groups should be prioritized for allocation? We designed a model that can optimally assign these resources.

SA58

Summit - Terrace Suite 2

Severe Trauma

Invited Session

Health Applications Society

Chair: Dimitris Bertsimas, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Vasiliki Stoumpou, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States

1 - Treatment of hip fragility fractures

Lisa Everest, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, United States, Catherine Ning

We develop machine learning models trained on multimodal sources of data to predict hip fragility fractures at the Hartford HealthCare (HHC) hospitals in Connecticut. In particular, we leverage tabular data, timeseries data, vision data, and language data to predict three main targets as identified by HHC orthopedic surgeons - readmission after surgery, post-operative complications within 3 days of surgery, and time to surgery after admission.

2 - Using Holistic AI to Diagnose AAST Injury Grade for Blunt Splenic and Liver injuries on CT scan and Direct Optimal Treatment

Carol Gao, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, Ikram Chairi, Muhammad Maaz

Liver injury is a critical concern in trauma care. As management depends on the severity of liver injury, cases are graded according to the American Association for the Surgery of Trauma (AAST) classification on a scale from I to V. However, recent literature has shown that CT reads by humans can misclassify liver injuries, leading to inappropriate management. We introduce a multimodal approach combining CT scans, as 3D images, unstructured clinical notes, vitals, and past medical history to predict the grading of liver injury for trauma patients. We then build a prescriptive model using patient management and outcome data to recommend operative versus non-operative approaches to patients with liver or spleen injuries. We apply our approach to n=3,331 trauma patients at Hartford Hospital in Connecticut to demonstrate the accuracy of our method.

3 - Prediction of Chronic Subdural Hematoma

Rohan Kumar, Boston University, Boston, MA, United States

A subdural hematoma is a type of intracranial bleed that occurs on the surface of the brain. Aging, trauma in the head, use of anticoagulants and alcohol, and coagulopathies have been associated with the occurrence of subdural hemorrhages. Subacute/chronic subdural hematoma, the main focus of this study, is expected to emerge as a predominant neurosurgical concern in the United States by 2030, primarily attributed to increasing life expectancy and the expanding utilization of antiplatelet and anticoagulation medications. The present work evaluates the efficacy of machine learning (ML) approaches used to predict the existence of chronic subdural hematoma. We perform these tasks using demographics, labs, medications and image data, by extracting embeddings from images, combining them with the tabular data and then training an XGBoost model to learn each task. We demonstrate that the combination of modalities is beneficial in successfully performing both tasks.

4 - OPOM 2.0: Continuity and Accuracy for an Optimized Prediction of Mortality for Candidates Awaiting Liver Transplantation

Maura Hegarty, Massachusetts Institute of Technology, Cambridge, MA, United States, Theodore Papalexopolous, Dimitris Bertsimas, Nikolaos Trichakis

In 2019, an Optimized Prediction of Mortality (OPOM) model was developed to predict the probability that a candidate awaiting liver transplant dies on the waitlist or is removed for being too sick within 90 days. OPOM, a machine-learning optimal classification tree-based model, demonstrated improved performance and equity in simulation and AUC over the then-current policy MELDNa which similarly predicts mortality in 90 days. However, OPOM predictions were found to be very sensitive to noise added to the data given its discrete decision-making process. As certain lab measurements are prone to noise, in particular creatinine and bilirubin, this is a significant drawback of OPOM. In 2023, the new and improved MELD3.0 was introduced and implemented into policy. MELD3.0 improves upon the performance and equity of MELDNa, in particular, for sex. An area of weakness of the MELD3.0 model is its poor predictive performance among patients diagnosed with hepatocellular carcinoma (HCC) which accounts for about 20-25% of patients waiting for liver donation each year. To address the issues resulting from the discontinuous nature of OPOM and the weak performance of MELD3.0 among HCC patients, we introduce OPOM 2.0 which employs the novel machine-learning method called deep trees. As deep trees are continuous, our predictions are more robust in the face of noise. We also observe enhanced overall model performance and most notably, for HCC patients.

SA60

Summit - Ballroom 2

Optimizing Last Mile Efficiency at Amazon

Invited Session

The Practice Section of INFORMS

Chair: Julie Poullet, Amazon - Last Mile, Seattle, WA, United States

1 - GPU based Route Planning with Cuopt

Bhargav Kunkulagunta, Amazon, Austin, TX, United States

Cuopt is a GPU based route solver, we have done benchmark to identify the potential oppurtunity with GPU based route solver and building a GPU based last mile stack for a potential GPU based GIS analytics system.

2 - Energy Aware Last-Mile Route Planning

André Snoeck, Amazon, Luxembourg, Luxembourg, Aniruddha Bhargava, Daniel Merchan, Mahmood Zangui

Amazon co-founded The Climate Pledge in 2019. As part of that Pledge, Amazon also announced to bring 100,000 electric delivery vehicles on the road by 2030. This goal presents challenges across multiple planning spaces including but not limited how to optimize the deployment of the EV fleet and how to plan EV routes. In this talk, we address the problem of predicting and incorporating energy estimates into last-mile planning systems. We show the need to move away from thinking about range and we propose that we should use energy as the basic unit of analysis for battery electric vehicles. Furthermore, we illustrate how incorporating EVs at large scale requires energy-aware planning throughout all decision making processes.

3 - Route time estimate optimization with non-linear opportunity cost awareness

Ivan Tereshchenko, Amazon, San Francisco, CA, United States

Last mile delivery services leverage machine learning models that predict delivery route duration at multiple stages of the route planning process. In this work we demonstrate the methodology for determining the duration of planned routes that uses quantile regression estimates and the newsvendor model with a generalized underage and overage cost functions. We derive approximate cost functions that reflect important business objectives. We show how our methodology can be used to simultaneously minimize late deliveries, driver dissatisfaction and route underutilization in ultra-fast delivery operations.

4 - Handling multi-objective contiguous unit allocation problems through Resources in a graph modeling

Julie Poullet, Amazon - Last Mile, Seattle, WA, United States, Andre Snoeck

In Amazon Logistics (AMZL), most of our deliveries are made by Delivery Service Partners (DSPs) which are independent local businesses. Typically, each Last Mile node is covered by a couple of DSPs operating on a daily basis. DSPs are assigned to an area in which they are operating the majority of their deliveries. Optimally assigning such areas to DSPs is a complex problem incorporating many dimensions, e.g., affinity and complexity, that are subject to change as the business evolves. We introduce a flexible and scalable graph-based approach that models these dimensions as resources in a Resource Constrained Shortest Path (RCSP) framework. We propose a Column Generation approach to solve the graph model, whose pricing problem is modeled as RCSP problem. We introduce a novel and extended A* algorithm leveraging these resources to discretize bounds and promising paths, reducing the number of enumerations by 96% and the run-time by 62% in complex instances. To demonstrate the modeling flexibility and business relevance of the approach, we optimize area assignments for the feasibility, fairness, and efficiency dimensions.

SA61

Summit - Ballroom 3

Cutting-Edge Research in Federated Learning

Invited Session

OPT: Machine Learning

Chair: Chhavi Sharma, Southern Methodist University, Dallas, TX

1 - Federated Learning with Data Heterogeneity

Jingrui He, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Federated learning allows multiple clients to collaborate in training a machine learning model with the orchestration of a central server, without requiring them to share the raw data. In real applications, it is usually the case that clients exhibit various types of data heterogeneity. In this talk, I will introduce some of our recent works modeling such data heterogeneity, including (1) clustered federated learning, which explores the optimal collaboration structure among clients; (2) personalized federated learning, aiming to generalize the global model to both participating and unparticipating clients; (3) robust federated learning against Byzantine clients. Towards the end, I will also share my thoughts regarding future directions.

2 - Communication-Efficient Federated Group Distributionally Robust Optimization

Zhishuai Guo, Northern Illinois University, DeKalb, IL, United States, Tianbao Yang

Federated learning faces significant challenges due to the heterogeneity in data volume and distribution across different clients. This heterogeneity often undermines the model's ability to generalize effectively across various data distributions including unseen ones. Previous methods to tackle this issue, particularly those based on group distributionally robust optimization (GDRO), tend to incur high communication and/or computational costs. To this end, our work introduces a suite of novel algorithms designed for communication-efficient Federated Group Distributionally Robust Optimization (FGDRO). Our contributions are threefold: Firstly, we introduce the FGDRO-CVaR algorithm, which optimizes the CVaR-constrained GDRO loss that focuses on the top-K clients with highest losses. Secondly, we present the FGDRO-KL algorithm, which optimizes a KL-regularized weighted loss. Lastly, we introduce FGDRO-KL-Adam that uses Adam updates local steps. Our theoretical results demonstrate that our methods achieve low communication and computational costs. We have validated the effectiveness of our algorithms through experiments in natural language processing and computer vision.

3 - Communication-Efficient Algorithms for Federated Bilevel Optimization

Junyi Li, University of Maryland, College Park, College Park, MD, United States

Bilevel Optimization has witnessed notable progress recently with new emerging efficient algorithms. However, its application in the Federated Learning setting remains relatively underexplored, and the impact of Federated Learning's inherent challenges on the convergence of bilevel algorithms remain obscure. In this talk, we investigate Federated Bilevel Optimization problems and propose a communication-efficient algorithm, named FedBiOAcc. The algorithm leverages an efficient estimation of the hyper-gradient in the distributed setting and utilizes the momentum-based variance-reduction acceleration. Remarkably, FedBiOAcc achieves the near-optimal sample and communication complexity and the linear speed up with respect to the number of clients.

4 - Multi-agent Sequential Decision Problems under Uncertainty

Chhavi Sharma, Southern Methodist University, Dallas, TX, United States, Harsha Gangammanavar

Sequential decision problems under uncertainty are ubiquitous and arise in various applications such as healthcare, energy infrastructure, and supply chains. Several efforts have been made to solve these problems in a distributed computing environment tailored to specific applications. However, a comprehensive tool for addressing multi-agent sequential decision problems under uncertainty remains elusive. This talk will delve into a particular class of such problems: multi-agent two-stage stochastic programming problems in which all agents have their own private two-stage stochastic program and are connected through a graph. We will present solution methods and analyze the behavior of the solutions obtained from them. We will also highlight the challenges when agents collaborate to approximate the recourse function.

SA62

Summit - Signature Room

Nanoretail Operations in Developing Markets

Invited Session

TutORial

1 - Nanoretail Operations in Developing Markets

Jan Fransoo, Tilburg University, Tilburg, Netherlands, Rafael Escamilla, Jiwen Ge

Across much of the developing world, family-operated nanostores provide daily grocery needs to billions of poorly paid consumers. This highly fragmented retail channel is of critical importance to consumer brands as in many markets this is the largest retail channel. We characterize the empirical context in which these stores operate, as well as the intricate operations that manufacturers and distributors put in place to supply them with their goods. We then elaborate on modeling operations execution and operations strategies to expose critical tradeoffs that are distinct from those in organized retail in developed markets. We discuss research results that have demonstrated why many manufacturers choose to serve this market directly and at high frequency, why manufacturers invest considerable effort in deploying sales agents networks, how this channel manages to remain competitive with modern organized retail such as convenience store chains, and how digitization and novel financing solutions provide a further competitive advantage to the nanoretail channel. Finally, we discuss how to conduct research in this retail segment and provide examples of novel contexts and business models that may open up new areas of nanoretail research.

SA63

Regency - 601

Technological Transformation: AI's Role in Shaping Future Societies

Invited Session

Information Systems

Chair: Che-Wei Liu, Arizona State University, Tempe, AZ, United States

1 - When AI Showcases: Comparing Human and Ai Streamer Effectiveness in Live E-Commerce

Meixian Wang, Temple University, Philadelphia, PA, United States, Guohou Shan, Keran Zhao, Jason Thatcher

Live Streaming E-commerce (LSE) is a synchronic format that embeds live streaming into e-commerce, where streamers sell products and interact with viewers synchronously. Many stores have launched LSE channels to attract traffic and increase sales. As competition for effective human streamers is intense, platform owners have developed artificial intelligence (AI) streamers as an alternative. However, it is unclear whether human or AI streamers are more effective at engaging viewers and selling products. Drawing on naturalness theory, we develop a research model that examines the differences between human and AI streamers' sales competency. Using collected data from Taobao Live, we found that human and AI streamers have different sales efficiency in different product categories (i.e., search and experience products). To further investigate the underlying mechanism, we conducted an experiment to link the roles of human and AI streamers, their different sales efficiency, and product categories. Our study provides insights for platforms and store owners seeking to utilize AI technology better to sell products and for designers interested in developing more effective AI.

2 - From Seeing to Signing up: the Co-Creative Dance of Humans and AI in Cultivating Symbiotic Uniqueness for Enhanced User Engagement

Che-Wei Liu, Arizona State University, Tempe, AZ, United States, Shenyang Jiang, Jiang Duan

The advent of Generative Artificial Intelligence (AI) heralds a new epoch in which users are not mere consumers but active participants in the creation of digital content. This evolution prompts an inquiry into the distinction between generative AI and its traditional counterparts. We contend that the essence of this difference is encapsulated in the concept of "Symbiotic Uniqueness," which we introduce to underscore the dynamic interplay between human input and the AI's ability to generate unique and random outputs. We argue that the two pillars of this concept are the effectiveness of the AI in meeting user intentions and the inherent randomness that fuels the creation of one-of-a-kind results. We devised a field experiment with a dual-axis framework: content visibility and loss framing. Our findings reveal that displaying AI-generated content upfront significantly boosts user registration. Moreover, a mixed strategy that combines both revelation and concealment of content suggests potential advantages, as it appears to strike a delicate balance between effectiveness and intrigue, slightly enhancing user registration rates. The application of loss framing, which plays on the randomness of AI-generated content, markedly amplified sign-up rates by tapping into the users' fear of missing out on unique creations. Surprisingly, the combination of content visibility and loss framing appeared to dilute the efficacies, pointing to a possible substitution effect. Lastly, we showcase that both content visibility and loss framing would be ineffective without active user participation in the co-creation process. This highlights the critical importance of symbiotic uniqueness in promoting user engagement.

3 - Multimodal Co-Attention Transformer for Video-Based Personality Understanding

Mingwei Sun, University of Maryland, College Park, MD, United States, Kunpeng Zhang

Video has emerged as a pervasive medium for communication, entertainment, and information sharing. With the consumption of video content continuing to increase rapidly, understanding the impact of visual narratives on personality has become a crucial area of research. While text-based personality understanding has been extensively studied in the literature, video-based personality prediction remains relatively under-explored. Existing approaches to video-based personality prediction can be broadly categorized into two directions: learning a joint representation of audio and visual information using fully-connected feed-forward networks, and separating a video into its individual modalities (text, image, and audio), training each modality independently, and then ensembling the results for subsequent personality prediction. However, both approaches have notable limitations: ignoring complex interactions between visual and audio components, or considering all three modalities but not in a joint manner. Furthermore, all methods require high computational costs as they require high-resolution images to train. In this paper, we propose a novel Multimodal Co-attention Transformer neural network for video-based affect prediction. Our approach simultaneously models audio, visual, and text representations, as well as their inter-relations, to achieve accurate and efficient predictions. We demonstrate the effectiveness of our method via extensive experiments on a real-world dataset: First

Impressions. Our results show that the proposed model outperforms state-of-the-art approaches while maintaining high computational efficiency. In addition to our performance evaluation, we also conduct interpretability analyses to investigate the contribution across different levels. Our findings reveal valuable insights into personality predictions.

4 - Group-Sparse Matrix Factorization for Transfer Learning of Word Embeddings

Kan Xu, University of Pennsylvania, Philadelphia, PA, United States, Xuanyi Zhao, Hamsa Bastani, Osbert Bastani

Unstructured text provides decision-makers with a rich data source in many domains, ranging from product reviews in retail to nursing notes in healthcare. To leverage this information, words are typically translated into word embeddings---vectors that encode the semantic relationships between words---through unsupervised learning algorithms such as matrix factorization. However, learning word embeddings from new domains with limited training data can be challenging, because the meaning/usage may be different in the new domain, e.g., the word “positive" typically has positive sentiment, but often has negative sentiment in medical notes since it may imply that a patient tested positive for a disease. In practice, we expect that only a small number of domain-specific words may have new meanings. We propose an intuitive two-stage estimator that exploits this structure via a group-sparse penalty to efficiently transfer learn domain-specific word embeddings by combining large-scale text corpora (such as Wikipedia) with limited domain-specific text data. We bound the generalization error of our transfer learning estimator, proving that it can achieve high accuracy with substantially less domain-specific data when only a small number of embeddings are altered between domains. Furthermore, we prove that all local minima identified by our nonconvex objective function are statistically indistinguishable from the global minimum under standard regularization conditions, implying that our estimator can be computed efficiently. Our results provide the first bounds on group-sparse matrix factorization, which may be of independent interest. We empirically evaluate our approach compared to state-of-the-art fine-tuning heuristics from natural language processing.

SA64

Regency - 602

Unpacking Inequalities in Social Media Algorithms and UGC

Invited Session

Social Media Analytics

Chair: Xi Wang, Central University of Finance and Economics, Beijing, N/A

1 - Impact of Perceived Personalized Recommendation on Content Novelty and Creative Performance on Content Creation Platforms

Xiji Zhu, Hitotsubashi University, Japan, Tokyo, Japan, Yuhan Zuo, Yu Yang, Junjie Wei

As content creation platforms proliferate and reshape the landscape of creative activities, the role of personalized recommendation in creative activities comes into focus. This study aims to understand how content creators’ perceived personalized recommendation influence their creative strategies and, in turn, their creative performance. We first develop and validate tools to measure perceived personalized recommendation through two samples (n1 = 209, n2 = 388), then use survey data of original video bloggers (n = 195) on a Chinese content creation platform to test our hypotheses. Based on signaling theory, we show that content creators’ perceived personalized recommendation promotes their content novelty and boost creative performance. We also find that these effects are intensified by content creators’ platform embeddedness. Our research advances the current studies on platform algorithmic control by revealing the consequences of an under-investigated algorithmic control form—algorithmic recommendation, and demonstrates the positive side of platform algorithmic control. We also diverge from the existing user-centric studies on platform algorithms and provide preliminary empirical evidence on how content creators’ responses to platforms’ personalized recommendation algorithms influence their subsequent creative outputs.

2 - Does AI Reduce Workplace Gender Inequality?

Pedro Masi, University of Tennessee, Knoxville, Knoxville, TN, United States, Yuanyang Liu, ChuanRen Liu

Many occupations are predominantly occupied by male or female workers. An important reason for such gender inequality in the workforce is the nature of certain tasks within these jobs, which tend to align more closely with either the capabilities or societal expectations of one gender over the other. For example, males are disproportionately represented in industry work while females are in pre-school education.

The purpose of this study is to determine if AI promotes gender diversity in the work-place. Specifically, this study examines whether exposure to AI reduces barriers to occupations traditionally dominated by a certain gender group. In essence, the study tests for the so called "deskilling effect" of AI technology. That is, if AI can perform certain tasks that were historically more challenging for a certain gender group to accomplish, then AI has effectively lowered the barrier of entry for that group. If so, then AI would allow that group to set foot into previously uncharted territory, thus, increasing gender diversity in the workplace.

3 - Analyzing the Dynamics and Dissemination of Gender Inequality on Weibo

Tiantian Liang, Central University of Finance and Economics, Beijing, China, People's Republic of

This study analyzes the dynamics and dissemination of gender inequality on Weibo through the lens of several key phrases, including "unmarried women," "older single women," "single motherhood," and "unmarried motherhood." We examine the differing perspectives on these topics among Chinese individuals to understand how these views diverge and spread across the platform. Our findings reveal significant variances in opinion and highlight the mechanisms through which these viewpoints are disseminated. This research offers valuable insights into social media dynamics and public opinion, with implications for understanding and managing cultural discourse.

4 - Gender Disparity in Expressed Emotions Within Online Health Communities

Yuehua Zhao, Nanjing University, Nanjing, China, People's Republic of

Online support groups offer a new way to users to communicate with others regarding certain health issues. Taking autism-related support groups on Facebook as an example, we examine whether the expressed emotions differ between female and male users in online health-related support groups and whether such gender disparity varied based on the topics of the groups. Experimental results reveal a significant gender difference of expressed emotions in the groups. We find that female users tended to express more positive emotions in the group discussions than the male group members did. In addition, users appeared to express different sentiments within the groups focused on various topics. Male users tend to convey more negative emotions in the group that related to treatment, while female users were more positive when posted in the research-related group than male users were. This study is beneficial for tracking and moderating the emotional environment in online support groups.

SA65

Regency - 603

NSF Program Director Panel

Panel Session

NSF

Co-Chair: Georgia-Ann Klutke, National Science Foundation, Alexandria, VA, United States

1 - Moderator Panelist

Georgia-Ann Klutke, National Science Foundation, Alexandria, VA, United States

2 - Panelist

Reha Uzsoy, North Carolina State University, Raleigh, NC, United States

3 - Panelist

Siqian Shen, University of Michigan, Ann Arbor, MI, United States

4 - Panelist

Linkan Bian, National Science Foundation, Alexandria, VA, United States

5 - Panelist

Alexandra Medina-Borjia, National Science Foundation, Alexandria, VA, United States

SA66

Regency - 604

AI-augmented Dynamic Decisions in Business

Invited Session

Artificial Intelligence

Chair: Elynn Chen, New York University, New York, NY, 10012, United States

1 - Doubly High-Dimensional Contextual Bandits: An Interpretable Model for Joint Assortment-Pricing

Linda Zhao, University of Pennsylvania, Philadelphia, PA, United States

Key challenges in running a retail business include how to select products to present to consumers (the assortment problem), and how to price products (the pricing problem) to maximize revenue or profit. Instead of considering these problems in isolation, we propose a joint approach to assortment-pricing based on contextual bandits. Our model is doubly high-dimensional, in that both context vectors and actions are allowed to take values in high-dimensional spaces. In order to circumvent the curse of dimensionality, we propose a simple yet flexible model that captures the interactions between covariates and actions via a (near) low-rank representation matrix. The resulting class of models is reasonably expressive while remaining interpretable through latent factors, and includes various structured linear bandit and pricing models as particular cases. We propose a computationally tractable procedure that combines an exploration/exploitation protocol with an efficient low-rank matrix estimator, and we prove bounds on its regret. Simulation results show that this method has lower regret than state-of-the-art methods applied to various standard bandit and pricing models. Real-world case studies on the assortment-pricing problem, from an industry-leading instant noodles company to an emerging beauty start-up, underscore the gains achievable using our method. In each case, we show at least three-fold gains in revenue or profit by our bandit method, as well as the interpretability of the latent factor models that are learned.

2 - New Optimization Models for AI Research

Tianyi Lin, Columbia University, New York, NY, United States

In this talk, we present several new results on minimizing a nonsmooth and nonconvex function under a Lipschitz condition. Recent works show that while the classical notion of Clarke stationarity is computationally intractable up to some sufficiently small constant tolerance, the randomized first-order algorithms find an approximate Goldstein stationary point with the finite-time convergence guarantee, which is independent of problem dimension. First of all, we present several motivating examples and explain why the notion of Goldstein stationary is important to the finite-time convergence guarantee. Second, we show that the randomization is necessary to obtain a dimension-independent guarantee, by proving a lower bound of $\Omega(d)$ for any deterministic algorithm that has access to both first-order and zeroth-order oracles. We also show that the zeroth-order oracle is essential to obtain a finite-time convergence guarantee, by showing that any deterministic algorithm with only the first-order oracle can not find an approximate Goldstein stationary point within a finite number of iterations up to sufficiently small constant parameter and tolerance. Finally, we establish the relationship between the notion of Goldstein stationary and the uniform smoothing, and propose several new zeroth-order methods with solid finite-time convergence guarantee.

3 - Dual Active Learning for Reinforcement Learning from Human Feedback

Will Wei Sun, Purdue University, West Lafayette, IN, United States, Pangpang Liu

Aligning large language models (LLMs) with human preferences is critical to recent advances in generative artificial intelligence. Reinforcement learning from human feedback (RLHF) is widely applied to achieve this objective. A key step in RLHF is to learn the reward function from human feedback. However, human feedback is costly and time-consuming, making it essential to collect high-quality conversation data for human teachers to label. Additionally, different human teachers have different levels of expertise. It is thus critical to query the most appropriate teacher for their opinions. In this paper, we use offline reinforcement learning (RL) to formulate the alignment problem. Motivated by the idea of D-optimal design, we first propose a dual active reward learning algorithm for the simultaneous selection of conversations and teachers. Next, we apply pessimistic RL to solve the alignment problem, based on the learned reward estimator. Theoretically, we show that the reward estimator obtained through our proposed adaptive selection strategy achieves minimal generalized variance asymptotically, and prove that the sub-optimality of our pessimistic policy. Through simulations and experiments on LLMs, we demonstrate the effectiveness of our algorithm and its superiority over state-of-the-arts.

SA67

Regency - 605

Case Competition

Award Session

Case Competition

Co-Chair: Matthew Drake, Duquesne University, Pittsburgh, PA, United States

1 - LandMover Equipment Company: Optimizing Servicizing Product Contract

Saurabh Bansal, Penn State, State College, PA, United States, Dan Guide, Wei Wu, Yinshi Gao, Yinshi (Alice) Gao

LMC is an original equipment manufacturer that sells capital-intensive earth-moving equipment. Recently, LMC decided to move towards a servicizing product model, which it expects to be more profitable and sustainable. Customers receive a product and return it back after use. The customers pay a fixed fee and a variable fee based on product usage. LMC performs extensive remanufacturing on the used product, e.g., taking the machine apart and repair. The remanufacturing cost depends on the customer's usage of the machine. LMC needs to determine the optimal contract for this business model.

2 - Potty Parity: Stadium Restroom Design

Vahid Roshanaei, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Setareh Farajollahzadeh, Ming Hu

Addressing long restroom wait times of women and LGBTQ+ individuals, this case presents tools to evaluate wait time disparities in various restroom setups. A stadium manager aims to enhance efficiency (wait time) and fairness (Totalitarian and Rawlsian Scores) among diverse gender preferences. In three modules, students (1) evaluate queuing parameters for mixed populations, (2) analyze wait-time disparities in restroom layouts, and (3) assess fairness from a DEI perspective. This case helps students understand queuing theory, efficiency-fairness trade-offs, and multi-objective problem-solving.

3 - Pricing Analytics for Porsche Drive Vehicle Subscription Program

Varun Gupta, University of North Georgia, Dahlonega, GA, United States, Vaughan Griffiths, Jing Li, Daniel Manion, Xiao Zhang

This case examines the pricing challenges of a luxury car subscription service: Porsche Drive, amidst a volatile automotive market. Facing unpredictable depreciation rates, a new intern analyzes cost-based, market-based, and value-based pricing models. Customers exhibit a heterogenous willingness to pay for the service, emphasizing the need for segmentation and reevaluation of pricing strategies. This case study explores the complexities of pricing in a rapidly changing industry, focusing on customer segmentation, market dynamics, and adapting pricing models to maximize subscription profit.

SA68

Regency - 606

Doing Good with Good OR

Award Session

Doing Good with Good OR

Chair: Amir Karimi, The University of Texas at San Antonio, San Antonio, TX, United States

Co-Chair: John Harris, NA, Tyrone, GA, United States

1 - Optimizing the Path Towards Plastic-Free Oceans

Baizhi Song, London Business School, London, United Kingdom

Rising ocean plastic pollution has considerable adverse effects on Earth and humans. We partner with a non-profit organization and use optimization to help clean up oceans from plastic faster. Specifically, we optimize the route of their plastic collection system in the ocean to maximize the quantity of plastic collected over time. The algorithm considers complex plastic dynamics and relevant operational constraints. We validate our algorithm on historical ocean data and achieve over 60% improvement over their current method. Our algorithm also informs the design of their new system.

2 - Strengthening Blood Transfusion System in Kenya by Analytic Tools

Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States

The PITS-Kenya project has advanced Kenya's blood transfusion services by an integrated OR tool set. We developed the first comprehensive process map and the first large-scale ‘vein-to-vein’ discrete event simulation platform to visualize and interact with the system. These innovations have streamlined operations, evaluated interventions, and are supporting clinical trials in 13 hospitals, impacting over 30,000 patients. Additionally, we are developing a novel blood drive planning tool to boost donor engagement, aiming to increase the annual blood supply from 30% to 50% of the target.

3 - Identification of Hazard Clusters for Priority Landmine Clearance as a Quadratic Knapsack Problem

Mateo Dulce Rubio, Carnegie Mellon University, Pittsburgh, PA, United States

Landmines pose a threat to war-affected communities even after conflicts have ended, in part due to the labor-intensive nature of demining. In this work we propose RELand, in partnership with the UN Mine Action Service, to identify hazard clusters for priority mine clearance under geographic and budget constraints using a novel QKP formulation. Ongoing field testing in actual demining operations in Colombia has so far identified three landmines in one prioritized area. The proposed system provides practical inputs for humanitarian demining with the potential for global application and impact.

4 - Optimizing Health Supply Chains with Decision-Aware Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States

We address the problem of allocating limited medical resources in a developing country by combining ML (to predict demand) with optimization (to optimize allocations). A key challenge is aligning the ML model's loss function with the decision loss in optimization. We propose a scalable decision-aware learning framework and successfully deployed in collaboration with the Sierra Leone government across 1,123 healthcare facilities nationwide, leading to a 15-29% increase in medicine consumption and improved real-world patient access to care.

SA69

Regency - 607

Human Engagement in AI and Digital Platforms

Invited Session

eBusiness

Chair: Jing Tian, Penn State University, State College, United States

1 - Can Hiding (Dis)like Help? An Investigation of the Impact of Performance Metrics Disclosure on Social Media Platforms

Yi Gao, Texas Tech University, Lubbock, TX, United States, Pei-yu Chen, Dengpan Liu

As social media becomes deeply integrated into daily life, concerns have emerged regarding the potential unintended negative effects on well-being due to social comparison and pressure. To address these concerns, platforms have started experimenting with various measures to safeguard the mental health of content creators. One approach is adjusting the disclosure of performance metrics. In this study, we investigate how changes in performance metric disclosures impact the platforms and creators by constructing a game-theoretic model that takes into account social comparison as well as the platform’s potential long-term loss from demand distortion driven by the disclosure of performance metrics. Our findings reveal that when the creators derive significant utilities or disutilities from social comparison, heightened disclosure can motivate less effective creators to increase their efforts. Interestingly, our analysis of the demand disparity between the two creators suggests that while disclosing more performance metrics may seemingly widen the popularity discrepancy between them, it does not necessarily lead to a greater disparity in demand. Moreover, we find that as the disclosure level increases, platforms may lower their commission rates and experience a decline in their profits. We also find that the optimal disclosure level decreases with the social comparison intensity and may sometimes increase with the platform’s traffic revenue per demand. These findings have significant implications for social media platforms on how to employ technologies for social good and regulators on how to respond to changes in the disclosure level.

2 - Learning for Approval

Xingchen Xu, University of Washington, Seattle, WA, United States, Yumei He, Yong Tan

In online matching platforms, users must exert effort (cost) and face the risk of rejection when seeking to forge connections with others. Beyond the mere absence of a match, there might exist an additional psychological cost borne by those who face rejection, distinct from the silent void left by an inactive counterpart. Concurrently, users exhibit heterogeneous preferences towards various participants on the opposing side of the platform, obfuscating the process of disentangling application costs, rejection costs, perceived probabilities of matching outcomes, and the intrinsic preferences that underpin user behavior. Acknowledging these empirical complexities, we construct a comprehensive Bayesian Learning model that jointly incorporates all the aforementioned mechanisms to characterize user behavior in matching markets. By leveraging a unique dataset from an online dating platform, we calibrate the model and thoroughly depict users' strategic behavior. On top of the estimation results, We develop counterfactual policies that harness the power of matching algorithms and large language models, thereby improving the efficiency of online matching process.

3 - Impact of the Representational Fidelity of AI Explanations and Perceived Bias of Ai Systems on Medical Decision-Making: Cognitive and Emotional Mechanisms

Rongen "Sophia" Zhang, Baylor University, Waco, TX, United States, Karl Werder, Kartikeya Negi, Balasubramaniam Ramesh

While Artificial intelligence (AI)-based decision-making systems have been shown to outperform them, medical experts are often reluctant to follow their recommendations because of the black-box nature of these systems. Explainable AI provides the promise of unpacking the black box of AI. Hence, this study aims to understand the impact of the representational fidelity of different explanations and the perceived bias of the AI system on the informed decisions of medical professionals. Drawing from effective use theory, we propose two mechanisms that mediate these effects: a cognitive learning mechanism and an emotional mechanism. We conducted two studies to test our hypotheses. In the first study, diabetes patients were asked to provide recommendations for their relatives based on their data, AI recommendations, and accompanying AI explanations. In the second study, healthcare providers, including physicians, nurse practitioners, and physician assistants, were asked to make decisions regarding the likelihood of developing sepsis based on patient data, AI recommendations, and various types of AI explanations. Our study provides a new perspective to the scholarly discourse on explainable AI and the effective use of AI by enhancing our understanding of how AI explanations influence medical decision-making and highlighting the importance of designing AI systems that support both the cognitive and emotional aspects of decision-making.

4 - Ground Truth Mechanisms in AI Development: A Conjoined Agency Perspective

Hengqi Tian, University of Colorado Denver, Denver, CO, United States, Arun Rai, Angelos Kostis, Jonny Holmström

Crafting ground truth data labels is instrumental but challenging in AI development. In contrast to the prevailing dominant objective view on ground truth labels and human-centered data labeling approaches, we adopt a conjoined agency perspective to theorize how the complementarities between humans and AI play out in organizing the data labeling process for AI development. We conceptualize ground truth data labeling as a highly iterative process involving reflection in action between human agency and AI agency. We propose that the level of ground truth uncertainty determines the composition of conjoined agency and the degree of reflection in action necessary to get the appropriate labels, which can lead to two different organizing principles emphasizing either accuracy or divergence. Our theoretical framework and propositions are expected to contribute to unpacking the composition and interactive dynamics of humans and AIs in constructing ground truth data labels and how learning occurs within human-AI interactions.

SA70

Regency - 701

Why I Love Applied Probability

Panel Session

Committee's Choice

Co-Chair: Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

1 - Moderator Panelist

Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

2 - Panelist

Amy Ward, The University of Chicago Booth School of Business, Chicago, IL, United States

3 - Panelist

Henry Lam, Columbia University, New York, NY, United States

4 - Panelist

Alan Scheller-Wolf, Tepper School of Business, Pittsburgh, PA, United States

SA71

Regency - 702

Statistical Learning via Generative Models, Simulation, and Optimization

Invited Session

Data Mining

Chair: Yao Xie, Georgia Institute of Technology, Atlanta, GA, United States

1 - Kernel Learning in Ridge Regression "Automatically" Yields Exact Low Rank Solution

Feng Ruan, Northwestern University, Evanston, IL, United States, Yunlu Chen, Yang Li, Keli Liu

This talk considers a variant of kernel ridge regression which simultaneously optimizes the prediction function and the reproducing kernel Hilbert space (RKHS) parameter. We identify a previously unnoticed phenomenon for this kernel learning objective where — under certain conditions — the global minimizer is exactly low rank with high probability. This phenomenon is interesting because the low rankness property is achieved without using any explicit regularization. Our theory makes correspondence between the observed phenomenon and the notion of low rank set identifiability from the optimization literature. The low rankness property of the finite sample solutions exists because the population kernel learning objective grows "sharply" when moving away from its minimizers in any direction perpendicular to the central mean subspace.

2 - Simulation Calibrated Machine Learning: Minimax Estimation via Hybrid Scientific and Statistical Models

Yiping Lu, Northwestern IEMS, Evanston, IL , IL, United States

Machine learning (ML) has achieved great success in a variety of applications suggesting a new way to build flexible, universal, and efficient approximators for complex high-dimensional data. These successes have inspired many researchers to apply ML to other scientific applications such as industrial engineering, scientific computing, and operational research, where similar challenges often occur. However, the luminous success of ML is overshadowed by persistent concerns that the mathematical theory of large-scale machine learning, especially deep learning, is still lacking and the trained ML predictor is always biased. In this talk, I'll introduce a novel framework of (S)imulation-(Ca)librated (S)cientific (M)achine (L)earning (SCaSML), which can leverage the structure of physical models to achieve the following goals: 1) make unbiased predictions even based on biased machine learning predictors; 2) beat the curse of dimensionality with an estimator suffers from it. The SCASML paradigm combines a (possibly) biased machine learning algorithm with a de-biasing step design using rigorous numerical analysis and stochastic simulation. Theoretically, I'll try to understand whether the SCaSML algorithms are optimal and what factors (e.g., smoothness, dimension, and boundness) determine the improvement of the convergence rate. Empirically, I'll introduce different estimators that enable unbiased and trustworthy estimation for physical quantities with a biased machine learning estimator. Applications include but are not limited to estimating the moment of a function, simulating high-dimensional stochastic processes, uncertainty quantification using bootstrap methods, and randomized linear algebra.

3 - Flow-Based Distributionally Robust Optimization

Chen Xu, Georgia Institute of Technology, Atlanta, GA, United States, Jonghyeok Lee, Xiuyuan Cheng, Yao Xie

We present a computationally efficient framework, called FlowDRO, for solving flow-based distributionally robust optimization (DRO) problems with Wasserstein uncertainty sets while aiming to find continuous worst-case distribution (also called the Least Favorable Distribution, LFD) and sample from it. The requirement for LFD to be continuous is
so that the algorithm can be scalable to problems with larger sample sizes and achieve better generalization capability for the induced robust algorithms. To tackle the computationally challenging infinitely dimensional optimization problem, we leverage flow-based models and continuous-time invertible transport maps between the data distribution
and the target distribution and develop a Wasserstein proximal gradient flow type algorithm. In theory, we establish the equivalence of the solution by optimal transport map to the original formulation, as well as the dual form of the problem through Wasserstein calculus and Brenier theorem. In practice, we parameterize the transport maps by a sequence of
neural networks progressively trained in blocks by gradient descent. We demonstrate its usage in adversarial learning, distributionally robust hypothesis testing, and a new mechanism for data-driven distribution perturbation differential privacy, where the proposed method gives strong empirical performance on high-dimensional real data.

4 - Spurious Stationarity and Hardness Results for Mirror Descent

Jiajin Li, University of British Columbia, Vancouver, BC, Canada

Despite the considerable success of Bregman proximal-type algorithms, such as mirror descent, in machine learning, a critical question remains: Can existing stationarity measures, often based on Bregman divergence, reliably distinguish between stationary and non-stationary points? In this paper, we present a groundbreaking finding: All existing stationarity measures necessarily imply the existence of spurious stationary points. We further establish an algorithmic independent hardness result: Bregman proximal-type algorithms are unable to escape from a spurious stationary point in finite steps when the initial point is unfavorable, even for convex problems. Our hardness result points out the inherent distinction between Euclidean and Bregman geometries, and introduces both fundamental theoretical and numerical challenges to both machine learning and optimization communities.

SA72

Regency - 703

Generative AI for Decision Making

Invited Session

Data Mining

Chair: Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Chong Liu, State University of New York at Albany, Albany, NY, 12222, United States

1 - Gradient Guidance for Diffusion Models: An Optimization Perspective

Hui Yuan, Princeton University, Princeton, NJ, United States, Yingqing Guo, Yukang Yang, Minshuo Chen, Mengdi Wang

Diffusion models have demonstrated empirical successes in various applications and can be adapted to task-specific needs via guidance. This work introduces a form of gradient guidance designed to adapt diffusion models towards user-specified optimization objectives. We study the theoretic aspects of a guided score-based sampling process, linking the gradient-guided diffusion model to first-order optimization. We show that adding gradient guidance to the sampling process of a pre-trained diffusion model is essentially equivalent to solving a regularized optimization problem, where the regularization term acts as a prior determined by pre-training data. Diffusion models are able to learn data's latent subspace, however, explicitly adding the gradient of an external objective function to the sample process would jeopardize the structure in generated samples. To remedy this issue, we consider a modified form of gradient guidance based on a forward prediction loss, which leverages the pre-trained score function to preserve the latent structure in generated samples. We further consider an iteratively fine-tuned version of gradient-guided diffusion where one can query gradients at newly generated data points and update the score network using new samples. This process mimics a first-order optimization iteration in expectation, for which we prove O(1/K) convergence rate to the global optimum when the objective function is concave.

2 - Stronger Together: Marrying Large Language Models with Traditional Symbolic Decision-Making

Yuandong Tian, Facebook, Menlo Park, CA, United States

While Large Language Models have achieved remarkable success, they are still weak for complicated tasks such as planning and optimization that look for optimal solutions for a given objective, subject to multiple constraints. In comparison, traditional symbolic solvers often give precise and guaranteed solutions to well-defined problems. In this talk, we explore a marriage of LLMs and symbolic solvers for better planning for hard optimization problems with natural language interface.

3 - From Minimax Score Matching to Optimal Sampling

Zehao Dou, Yale University, New Haven, CT, United States

Score-based diffusion models have significantly enhanced the performance of image, audio and video generation in the few past years. As a key ingredient, score estimation plays the most important role in the overall quality of diffusion models. In this work, we study the estimation of score function for the noisy probability distribution from n i.i.d observations. Assuming the Hölder smoothness and the compact support set of the original density, we establish the minimax rate of the mean integrated squared error for score estimation under all time steps. Without any additional logarithmic gaps, we perfectly match the upper and lower bounds, and achieve the optimal density estimation by diffusion models with the minimax score estimator plugged in.

4 - Representation-Based Reinforcement Learning

Bo Dai, Georgia Tech, ATLANTA, CA, United States

The majority reinforcement learning (RL) algorithms are largely categorized as model-free and model-based through whether a world model is learned in the algorithm. However, both of these two categories have their own issues, especially incorporating with function approximation: the exploration with arbitrary function approximation in model-free RL algorithms is difficult, while optimal planning becomes intractable in model-based RL algorithms with neural nonlinear world models.

In this talk, I will present our recent work on exploiting the power of representation in RL to bypass these difficulties, while enjoys best of both worlds. Specifically, we designed practical algorithms for extracting useful representations from world model, with the goal of improving statistical and computational efficiency in exploration vs. exploitation tradeoff and empirical performance in RL. We provide rigorous theoretical analysis of our algorithm, and demonstrate the practical superior performance over the existing state-of-the-art empirical algorithms on several benchmarks.

5 - OptiGuide: Language models for supply chain management (and beyond)

Ishai Menache, Microsoft, Redmond, WA, United States

We will provide an overview of our work on applying language model technology for supply chain management. We will also discuss future directions in the intersection of language models and operations.

SA73

Regency - 704

Waiting Time Prediction using Machine Learning

Invited Session

Data Mining

Chair: Yaron Shaposhnik, University of Rochester, Rochester, NY, United States

1 - Prediction of hospitalization and waiting time within 24 hours of emergency department patients with unstructured text data

Hyeram Seo, Asan Medical Center & Yonsei Univercity, (UTC+09:00) Asia/Seoul, Korea, Republic of

Overcrowding of emergency departments is a global concern, leading to numerous negative consequences. This study aimed to develop a useful and inexpensive tool derived from electronic medical records that supports clinical decision-making and can be easily utilized by emergency department physicians. We presented machine learning models that predicted the likelihood of hospitalizations within 24 hours and estimated waiting times. Moreover, we revealed the enhanced performance of these machine learning models compared to existing models by incorporating unstructured text data. Among several evaluated models, the extreme gradient boosting model that incorporated text data yielded the best performance. This model achieved an area under the receiver operating characteristic curve score of 0.922 and an area under the precision-recall curve score of 0.687. The mean absolute error revealed a difference of approximately 3 hours. Using this model, we classified the probability of patients not being admitted within 24 hours as Low, Medium, or High and identified important variables influencing this classification through explainable artificial intelligence. The model results are readily displayed on an electronic dashboard to support the decision-making of emergency department physicians and alleviate overcrowding, thereby resulting in socioeconomic benefits for medical facilities.

2 - Probabilistic Forecasting of Patient Waiting Times in an Emergency Department

Siddharth Arora, University of Oxford, Oxford, United Kingdom

This study aims to generate probability distribution forecasts of individual patient waiting times in an emergency department (ED). Wait time estimates can offer low-acuity patients and first responders a more comprehensive view of the possible waiting trajectory and provide reliable inputs to inform prescriptive modelling of ED operations. We define wait times as the time elapsed from patient registration at the ED to the start of treatment. Previous studies on modeling wait times have typically focused on generating a point estimate. However, waiting time is inherently uncertain, and the distribution is asymmetric. Thus, communicating only a point forecast can be uninformative and potentially misleading. This motivates us to generate probabilistic estimates of wait times. Using a large patient-level data set, we extract the following categories of predictor variables during the modeling: (1) calendar effects, (2) demographics, (3) staff count, (4) ED workload resulting from patient volumes, and (5) the severity of the patient condition. We demonstrate that publishing probabilistic waiting time estimates can inform patients and ambulance staff when selecting an ED from a network of EDs, leading to a more uniform spread of patient load across the network.

3 - Waiting Time Prediction using Machine Learning

Yashi Huang, University of Rochester, Rochester, NY, United States, Arik Senderovich, Yaron Shaposhnik

We develop a system for a large cancer institute for predicting in real-time patients' waiting times. The system consists of a collection of machine learning and simulation models that are assembled to make predictions (e.g., different models are used in different cases in different stages of the service process). The system is designed to overcome various challenges such as stochasticity (e.g., unexpected cancellations, and random service and arrival times), heterogeneity (e.g., of patients, providers, and workflow), complex service process (e.g., multiple interconnected stages), and limitations of the available data (e.g., noisy data and changes to the schedule that may not be electronically recorded). At the same time, the system is designed to provide the users explanations about why predictions are made and alert them about potentially inaccurate predictions and delays to the schedule. We report on the performance of our models for the various prediction tasks using a real-world data set.

SA74

Regency - 705

Optimization and Learning for Future Electricity Market Design

Invited Session

ENRE: Electricity

Chair: Rajni Bansal, Johns Hopkins University, Baltimore, MD, 92093, United States

Co-Chair: Patricia Hidalgo-Gonzalez, University of California, San Diego, La Jolla, CA, United States

1 - Balancing Fairness and Efficiency in Energy Resource Allocations

Baosen Zhang, University of Washington, SEATTLE, WA, United States, Jiayi Li, Matthew Motoki

Bringing fairness to energy resource allocation remains a challenge, due to the complexity of system structures and economic interdependencies among users and system operators' decision-making. The rise of distributed energy resources has introduced more diverse heterogeneous user groups, surpassing the capabilities of traditional efficiency-oriented allocation schemes. Without explicitly bringing fairness to user-system interaction, this disparity often leads to disproportionate payments for certain user groups due to their utility formats or group sizes. We addresses this challenge by formalizing the problem of fair energy resource allocation and introducing the framework for aggregators. This framework enables optimal fairness-efficiency trade-offs by selecting appropriate objectives in a principled way. By jointly optimizing over the total resources to allocate and individual allocations, our approach reveals optimized allocation schemes that lie on the Pareto front, balancing fairness and efficiency in resource allocation strategies.

2 - Locational Marginal Price in Electricity Markets

Steven Low, Caltech, Pasadena, CA, United States

In this brief tutorial, I will explain the basic economic dispatch problem in the electricity market and properties of locational marginal prices (LMPs). Specifically, I will present the DC optimal power flow problem that maximizes social welfare over real-time dispatch decisions subject to injection capacity and line flow limits. The resulting LMPs are nodal prices. They balance supply and demand, align individual incentives with social welfare, and are different at different nodes if and only if line congestion prices are nonzero. I will describe how injection limits and line congestion can cause LMPs at some nodes to be negative, but the system operator will never run cash negative. Time permitting, I will extend the basic economic dispatch problem to a security-constrained problem where dispatch decisions must be made under network contingencies.

3 - Fair Allocations of Cooperation Gains in Peer to Peer Electricity Markets

Shmuel Oren, UC Berkeley, Berkeley, CA, United States, Tomas Valencia Zuluaga

The advent of differentiated prices for consuming (buying) and injecting (selling) electricity at the residential level promotes the emergence of local peer-to peer electricity markets for prosumers, which can deliver savings to participants as long as an attractive cost sharing mechanism can be designed. Building on cooperative game theory models that have been proposed in the literature, we define the uniform price core, a class of desirable distribution of savings in this context, and prove constructively that it is not empty. We propose the shadow price imputation, a computationally efficient stable uniform price imputation, which we show to be equivalent to a dual imputation in the sense of cooperative linear production games. We then compare the shadow price imputation to other imputations in the uniform price core through the lens of fairness. To overcome the challenge of computational tractability for large numbers of participants, we employ a statistical sampling methodology for approximating the core and apply it to optimization problems devised to obtain a fair imputation. We present theoretical results and numerical experiments and examples illustrating our approach.

4 - Energy Storage Bidding Mechanism for Temporally-Coupled Electricity Markets

Rajni Bansal, University of California San Diego, San Diego, CA, United States, Patricia Hidalgo-Gonzalez

Electricity markets typically consist of two stages: a day-ahead (forward) market and a real-time (spot) market. This study proposes a market mechanism for a temporally coupled and stochastic two-stage multi-interval electricity market with different types of participants, such as generators and energy storage. Competitive equilibrium analysis is used to examine the impact of day-ahead decisions on the bidding curve of energy storage in the real-time market. Numerical experiments are conducted using New York ISO data to demonstrate the benefits of incorporating day-ahead decisions and uncertainty in the bidding function and to establish limits on the proposed market mechanisms.

SA75

Regency - 706

Net-Zero Emissions Energy Systems: Equitable Transitions

Invited Session

ENRE: Energy-Climate

Chair: Aleksander Grochowicz, University of Oslo, Oslo, Norway

Co-Chair: Jacqueline Dowling, Carnegie Science, 260 Panama St., Providence, 94305, United States

Co-Chair: Edgar Virgüez, Carnegie Institution for Science, Stanford, CA, United States

1 - Assessing multiple objectives in participatory energy modelling

Aleksander Grochowicz, University of Oslo, Oslo, Norway, Oskar Vågerö, Koen van Greevenbroek, Maximilian Roithner

As climate mitigation efforts focus strongly on decarbonising energy systems, some communities are forced to rapidly transform their fossil-based energy infrastructure.
In many places, the transition to renewable systems can be met with skepticism due to the top-down implementation of national targets.
In this case study, we consider the Arctic settlement Longyearbyen, which is the world's northernmost settlement at 78 degrees latitude.
Founded as a coal mining town, the Norwegian government has shut down its coal power plant and aims to transition to a renewable off-grid

system.

We use a novel participatory modelling approach to investigate the preferences of the local population regarding the future energy system in Longyearbyen.
Based on results from an open energy system optimisation model, we use near-optimal methods to describe many possible outcomes and additionally compare their performance under different metrics and objectives.
We develop a tool through which a significant part of the population interacts with the challenging tasks of weighing off different objectives such as e.g. costs, emission reduction, and vulnerability of energy supply.
We find that generally the local population is willing to accept some higher costs if other priorities can be achieved and acknowledge the trade-offs and challenges in the decision process.
By laying bare the complexities and offering very diverse system alternatives, our approach can bridge the communication and information gap between stakeholders, modellers, and policymakers.

2 - Resilience of Offshore Wind Technologies under Hurricane Landfalling Within the PJM Interconnection

Avery Barnett, Princeton University , Princeton , NJ, United States, Jesse Jenkins

New Jersey (NJ) is not spared from hurricane impacts and will have to work towards ensuring that social and technical impacts are mitigated and that infrastructure is sufficiently hardened. In line with the recent executive orders that mandate 100% clean energy by 2035 and 11,000MW of offshore wind by 2040, my research project focuses on the resilience of offshore wind technologies under hurricane landfall within PJM. Through varying the penetration of offshore wind technologies in the grid, I will investigate the options available to balance resilience and affordability, the impact of offshore wind on the overall system, and the level of electricity supply disruption at different hurricane intensities. By using GenX and an extreme weather model, the impacts of the electric grid within NJ will be assessed and solved for a range of different configurations, including varying offshore wind penetrations and other resilience factors, which will then be impacted by varying intensities of hurricanes where wind speed is used as a proxy. The impact of electricity supply disruption that results from the failure and how much failure is attributed to offshore wind within each configuration will be assessed. This research hopes to provide a first step in understanding the resilience of NJ, and by extension, the PJM grid against hurricanes. Future research can be informed through adaptation to consider other electricity grid threats such as storm surges, flooding, and similar hazards.

3 - Disparities in Per Capita Energy Consumption Within and Between Countries

Edgar Virguez, Carnegie Institution for Science, Stanford, CA, United States

Addressing issues related to equity and poverty requires a better understanding of the more than 1000-fold differences in per capita energy consumption worldwide. Here, we integrate national energy balances, population, income distribution, and gross domestic product data, to identify patterns in the current distribution of per capita energy consumption by quantifying existing disparities within and between countries. Among the profound disparities evident from this analysis, we find that 1% additional primary energy supply worldwide would be sufficient to double the energy available to the 10% of the most energy-impoverished people. In contrast, the 10% of the world's population with the highest energy consumption uses ~41% of the world's energy. The half of the people in the world using the least energy consume about the same amount of energy as 1.6 % of most profligate energy users. To bring the half of people who consume the least amount of energy to the worldwide median would increase global energy demand by 13.4%. Bringing the 1% of people who consume more energy than the 99th percentile level to the 99th percentile level would make enough energy available to bring the 31.1% of people who are most energy-impoverished to the level of the 31.1th percentile. Countries differ more in their mean per capita energy consumption than in the degree of within-country disparity. We conclude that relatively modest increases in global energy production (and/or redistribution of energy use) could produce large fractional increases in energy provision for the most energy-impoverished people in the world

4 - Climate change loss and damages (L&D) and reparations for equitable net-zero transitions.

Kiran Chawla, Stanford University, Palo Alto, CA, United States, Ines Azevedo

We use integrated assessment model based projections to understand what economic losses from climate change will be in different world regions. Even if the world optimally switches to a clean energy economy through global coordination, the impacts of climate change that are going to be baked into a warmer world need to be addressed through financial support from countries most responsible for emissions. We estimate the magnitude of these transfers and show how they are orders of magnitude greater than current international commitments, demonstrating a need to scale up ambition by countries for equity improving transfers.

5 - Air pollution inequity in net-zero targets in United States

Candelaria Bergero, UCI, Irvine, CA, United States

Meeting the target of the Paris Agreement of 1.5°C calls for large amounts of carbon dioxide removals (CDR). The U.S. net-zero GHG emissions target could require as much as 1.9 Gt of CDR by mid-century. Because GHGs and air pollution are commonly co-emitted, net-zero targes generally improve air quality and health outcomes. However, the magnitude and distribution of such health benefits will ultimately depend on the mitigation pathway chosen. Here we combine a series of models and analyze the pollution and health related impacts on local communities of two net-zero targets for the contiguous U.S. with varying amounts of CDR. We find that, while net-zero targets improve air quality and could avoid as much as 84,000 PM2.5-related deaths compared to a business-as-usual case, there are difference between an unrestricted CDR scenario and a restricted one in terms of mortality and its distribution across races and income groups. Meeting the dual goals of mitigating climate change and achieving a more equitable distribution of the benefits of the transition will require purposeful policy planning.

SA76

Regency - 707

Data-Driven, Human-Centered Decision Making in Stochastic Service Systems

Invited Session

MSOM: Service Operations

Chair: Yueyang Zhong, London Business School, London, N/A

1 - Learning to Defer in Content Moderation: the Human-AI Interplay

Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States, Wentao Weng

Ensuring successful content moderation is vital for a healthy online social platform where it is necessary to responsively remove harmful posts without jeopardizing non-harmful content. Due to the high-volume nature of online posts, human-only moderation is operationally challenging, and platforms often employ a human-AI collaboration approach. A typical machine-learning heuristic estimates the expected harmfulness of incoming posts and uses fixed thresholds to decide whether to remove the post (classification decision) and whether to send it for human review (admission decision). This can be inefficient as it disregards the uncertainty in the machine-learning estimation, the time-varying element of human review capacity and post arrivals, and the selective sampling in the dataset (humans only review posts filtered by the admission algorithm).

We introduce a model to capture the human-AI interplay in content moderation. The algorithm observes contextual information for incoming posts, makes classification and admission decisions, and schedules posts for human review. Non-admitted posts do not receive reviews (selective sampling) and admitted posts receive human reviews on their harmfulness. These reviews help educate the machine-learning algorithms but are delayed due to congestion in the human review system.

We propose a near-optimal learning algorithm that carefully balances the classification loss from a selectively sampled dataset, the idiosyncratic loss of non-reviewed posts, and the delay loss of having congestion in the human review system. To the best of our knowledge, this is the first result for online learning in contextual queueing systems and hence our analytical framework may be of independent interest.

2 - Ambiguous Dynamic Treatment Regimes: A Reinforcement Learning Approach

Soroush Saghafian, Harvard University, Cambridge, MA, United States

Researchers in different domains often want to use observational data and provide recommendations that can yield causal improvements. When using available methods, they often have to rely on assumptions that are violated in real-world applications (e.g., medical decision-making or public policy), especially when (a) the existence of unobserved confounders cannot be ignored, and (b) the unobserved confounders are time-varying (e.g., affected by previous actions). Motivated by a case study of finding superior treatment regimes for patients who underwent transplantation in our partner hospital (Mayo Clinic), we introduce a new methodology termed Ambiguous Dynamic Treatment Regimes (ADTRs), in which the causal impact of treatment regimes is evaluated based on a “cloud” of causal models. We then develop new reinforcement learning algorithms and run detailed experiments using real-world patient level data.

3 - Detecting Service Slowdown Using Observational Data

Gal Mendelson, Faculty of Data and Decision Sciences, Technion, Adi, Israel, Kuang Xu

In this talk, I will discuss the interplay between control and statistical analysis in service systems. The choice of the control mechanism is key for obtaining desired performance and fast, online detection of persistent changes in service speed is key to maintaining desired service quality and availability. I will demonstrate that the choice of control has a substantial impact on the nature of the data that is being generated and consequently changes how it should be used. I will discuss the limitations of using state data (e.g., queue lengths) for detection and present a new statistic, namely action data, and demonstrate that it can be very powerful.

4 - Integrating In-Person Specialists' Care with AI-Driven Telemedicine

Fanying Chen, Boston University, Boston, MA, United States, Opher Baron, Abraham Seidmann

Telemedicine has been slowly gaining popularity, and the COVID-19 pandemic dramatically boosted its usage. With an acute shortage of physicians, large healthcare organizations have been looking at the best ways to integrate virtual, digital, and condition-specific automated care programs to create a unified, comprehensive clinical journey for their members. Working with the leadership of a large HMO, we investigate the option of incorporating an independent third-party service that will offer select patients rapid access to virtual AI-driven dermatological diagnostic systems. We present a novel model of such a hybrid service system designed to shorten waits for this specialty medical care. Under this model, the automated appointment scheduling system dynamically activates the (far more expensive) virtual AI-driven care line only when the primary (in-person) service line becomes overloaded. We use extensive field data analytics to show that our proposed hybrid service-delivery model provides significant service improvements, even when the virtual AI-driven telecare is called to serve only a small proportion of patients.

SA77

Regency - 708

Logistics under Contested Conditions

Invited Session

Computing Society

Chair: Daniel Bienstock, Columbia University, New York, NY, United States

1 - Dreamcatcher-Operations

Jeff Linderoth, University of Wisconsin-Madison, Madison, WI, United States

Given friendly and adversary Orders of Battle and projected movements, which weapons should be assigned over time to missions to maximize the damage to high-priority targets? Dreamcatcher-Operations (DCOPS) is a mixed-integer programming-based tool for helping decision-makers answer this question.

We will discuss the design and implementation of DCOPS, including its newest features aimed at measuring and mitigating the risk of operations.

2 - Planning Logistics Under Adversarial Conditions

Blake Sisson, Columbia University, New York, NY, United States, Daniel Bienstock, Alexandra Newman, Justin Kilb, Caleb Fluker

We explore multiple mixed-integer programming- and machine-learning-based tools designed to enhance operational planning for contested logistics that rapidly evaluate the results of planning models, making them resilient to adversarial uncertainty and enabling real-time strategic adaptations. Acknowledging the high degree of uncertainty involved in contested logistics, we focus on an iterative algorithm that explores blue and red multi-period action plans. Building on the methods of Stackelberg competition, the process simulates sequences of actions and reactions, generating a large solution space that quantifies the implications of both rational and irrational decisions. By simulating a variety of possible outcomes, in which each iteration finds the best currently available set of decisions, the approach ensures that blue's strategy remains robust, effectively reducing vulnerability regardless of the unpredictable nature of red's actions. These tactical tools are designed to be fast-solving and user-friendly, allowing decision-makers without deep technical expertise to utilize advanced optimization and artificial intelligence capabilities.

3 - Layered Graph Security Games & Contested Logistics

Jakub Cerny, Columbia University, NY, NY, United States, Chun Kai Ling, Christian Kroer, Garud Iyengar

Security games model strategic interactions in adversarial real-world applications. Such applications often involve extremely large but highly structured strategy sets (e.g., selecting a distribution over all patrol routes in a given graph). In this paper, we represent each player's strategy space using a \textit{layered graph} whose paths represent an exponentially large strategy space. Our formulation entails not only classic pursuit-evasion games, but also other security games, such as those modeling anti-terrorism and logistical interdiction. We study two-player zero-sum games played on such graphs, and develop scalable algorithmic approaches for computing optimal plans for each player.

4 - A Fast Algorithm for Dual Decomposition in Vehicle Delivery Networks

Samuel Tan, Cornell University, Ithaca, NY, United States, Peter Frazier, Matthew Ford, Huseyin Topaloglu

We present a fast algorithm for solving dual decomposition in discrete-time vehicle delivery problems. We first use Lagrangian relaxation on a coupling constraint between vehicle movements and inventory levels to yield a dual decomposition, which can be solved with dual ascent. Moreover, the sets of problems can be further grouped into (1) a set of problems corresponding to individual vehicles, and (2) a set of problems corresponding to each item type and location. By appealing to the economic interpretation of the dual variables as prices in a marketplace, we show that the vehicle problems can be solved efficiently with dynamic programming, and that the item/location problems can be solved analytically in closed-form. We also provide an algorithm for generating primal feasible solutions in the case that the dual ascent terminates early. On a variety of realistic test instances, we show that our method yields objective values which are competitive with those produced by a mixed integer programming solver with significantly less computation.

SA78

Regency - 709

Advanced Computational Algorithms for Decision-making Process

Invited Session

Computing Society

Chair: Yingying Zhang, Western Michigan University, PORTAGE, MI, United States

1 - Unveiling Competitive Dynamics: Leveraging Generative Large Language Models for Textual Analysis of Firm Actions and Their Impact on Competitive Advantage

Yuanyuan Chen, University of Alabama Culverhouse College of Commerce, Tuscaloosa, AL, United States

In this study, we delve into the intricate relationship between manufacturers' competitive actions and their resulting competitive advantages. Our focus is on firms' strategic maneuvers in the domains of artificial intelligence and data analytics. Leveraging Generative Large Language Models (GLLMs) and GPT models, we embark on a comprehensive textual analysis of manufacturers' financial filings and press releases. This meticulous analysis enables us to gather pertinent data on firms' competitive actions, product development initiatives, and the phenomenon of servitization. Our investigation extends to assessing the impact of firms' actions in the area of AI and data analytics on key performance indicators that are crucial for business success, including customer relationship management, profitability, and stock market returns. Beyond the empirical findings, this study has significant managerial implications, shedding light on strategic decision-making processes and operational management practices within manufacturing firms. Furthermore, our methodological approach contributes to the field by pioneering an innovative use of GLLMs and GPT in textual mining of large data sets.

2 - Fair Conformal Prediction and Risk Control

Linjun Zhang, Rutgers University, New Brunswick, NJ, United States

Multi-calibration is a powerful and evolving concept originating in the field of algorithmic fairness. For a predictor f that estimates the outcome y given covariates x, and for a function class C, multi-calibration requires that the predictor $f(x)$ and outcome y are indistinguishable under the class of auditors in C. Fairness is captured by incorporating demographic subgroups into the class of functions C. Recent work has shown that, by enriching the class C to incorporate appropriate propensity re-weighting functions, multi-calibration also yields target-independent learning, wherein a model trained on a source domain performs well on unseen, future target domains(approximately) captured by the re-weightings. The multi-calibration notion is extended, and the power of an enriched class of mappings is explored. HappyMap, a generalization of multi-calibration, is proposed, which yields a wide range of new applications, including a new fairness notion for uncertainty quantification (conformal prediction), a novel technique for conformal prediction under covariate shift, and a different approach for fair risk control, while also yielding a unified understanding of several existing seemingly disparate algorithmic fairness notions and target-independent learning approaches. A single HappyMap meta-algorithm is given that captures all these results, together with a sufficiency condition for its success. Time permitting, the application of HappyMap to computer vision and large language models will also be discussed.

3 - Model-Based Clustering Analysis on the Spatial-Temporal and Intensity Pattern of Tornadoes

Rong Zheng, Western Illinois University, Macomb, IL, United States, Yana Melnykov, Yingying Zhang

Tornadoes are one of the nature's most violent windstorms that can occur all over the world. Many papers have been written to analyze tornadoes. However, the developed models fail to consider the dynamic nature of tornado paths and covariates that affect the tornadoes behaviors. In the proposed models, finite mixture models are employed to describe various tornado patterns, and the clustering analysis on the spatial-temporal and intensity patterns of the tornadoes are also conducted.

4 - On Regime Changes in Text Data Using Hidden Markov Model of Contaminated Vmf Distribution

Yingying Zhang, Western Michigan University, Portage, MI, United States, Shuchismita Sarkar, Yuanyuan Chen, Xuwen Zhu

This paper presents a novel methodology for analyzing temporal directional data with scatter and heavy tails. A hidden Markov model with contaminated von Mises-Fisher emission distribution is developed. The model is implemented using forward and backward selection approach that provides additional flexibility for contaminated as well as non-contaminated data. The utility of the method for finding homogeneous time blocks (regimes) is demonstrated two real-life text data sets containing presidential addresses and corporate financial statements respectively.

Sunday, October 20, 9:30 AM - 10:35 AM

Summit - Ballroom 1

How OR Helps the United Nations Deliver Food and Hope

Plenary/Keynote Session

Plenary

Chair: Stefan Karisch, Amazon

1 - How OR Helps the United Nations Deliver Food and Hope

Koen Peters, World Food Programme, Utrecht, Netherlands

After years of steady decline, global hunger is on the rise again. Over 309 million people are on the brink of starvation as conflicts escalate worldwide. This alarming increase in food insecurity is driven by the many concurrent challenges the world is facing: the slow post-COVID recovery of global markets, a rise in natural disasters, and deepening conflicts. The United Nations World Food Programme (WFP) has responded by placing Operations Research (OR) at the core of its emergency response, transforming humanitarian work. How can analytics be harnessed to combat the rising tide of global hunger? What does it take to apply these sophisticated tools in WFP’s complex operational environment? Through OR, the WFP has not only streamlined food distribution, cutting delivery times significantly, but also provided a beacon of hope. As we confront these challenges, the continued innovation in OR is not just promising—it is a critical step towards a world without hunger.

Sunday, October 20, 10:45 AM - 12:00 PM

SB01

Summit - 320

People Operations and Bias

Invited Session

Service Science

Chair: Lai Wei, Boston College, Chestnut Hill, MA, 02467, United States

Co-Chair: Teng Ye, University of Minnesota, Twin Cities, Minneapolis, MN, 55455, United States

1 - Design Multidimensional All-Pay Procurement Auction with Loss-Averse Bidders

Shan Li, City University of New York, New York, NY, United States, Xianghua Wu, Kay-Yut Chen

Procurement auctions involving multidimensional bids, typically consisting of a proposal and a price, are common in various industries. One notable aspect of these auctions is the all-pay component, where bidders must invest time and effort upfront to prepare a proposal, regardless of whether they win the contract. If a bidder loses the auction, the cost of developing the proposal is unrecoverable. Given that individuals tend to be loss averse, meaning they are more sensitive to losses than to gains of the same magnitude, loss aversion is particularly relevant in this setting, especially regarding the all-pay component. We analyze the impact of bidder loss aversion using game theoretical analysis and find that in the unique symmetric equilibrium of this setting, loss aversion consistently reduces equilibrium bid quality. Additionally, the effect on the buyer's equilibrium expected utility depends on the degree of loss aversion and how the buyer values the bidder's proposal. To mitigate the potential negative impact of loss aversion, we investigate two quality compensation policies commonly used in practice and studied in the literature: the proportional and the flat compensation policy. We show that both policies can improve social welfare without harming either side. Furthermore, we find that the flat compensation policy is preferable when the bidder population exhibits low levels of loss aversion.

2 - When the Personal is Professional: the Racialization of DEI Leadership

Rebecca Ponce de Leon, Columbia University, New York, NY, United States, James Carter

Spurred in large part by mega-events that laid bare the persistence of societal inequality, organizations have increasingly sought to demonstrate their commitment to diversity, equity, and inclusion (DEI) in recent years. As part of these efforts, many companies have created new leadership positions, like Chief Diversity Officer, to manage the ever-shifting diversity landscape. Despite this increase in interest, very little is known about the candidate attributes relied upon to select DEI leaders. The work current theorizes that these role are viewed as personal, identity-based roles, resulting in the reliance on personal attributes—like one's identity and background—over professional attributes—like one's work experience—to inform perceptions about candidates' fit and hireability. Further, decision makers use candidate race as a proxy to infer personal fit, resulting in general preferences for candidates of color over White candidates. We demonstrate these effects across six experiments, including studies exploring cognitions about the attributes necessary for success, selection decisions, and leader evaluations for DEI versus non-DEI leadership roles. We conclude by discussing critical implications for both theory and organizations.

3 - The Effect of Social Class Signals on Entry-Level Labor Market

Lai Wei, Boston College, Chestnut Hill, MA, United States, Teng Ye

Candidate evaluation in hiring and recruiting has drawn attention in the literature, specifically due to the prevalent biases within these processes. In this study, we focus on a different channel that could lead to evaluation bias in job interviews: social class. Understanding social class biases in the labor market is especially important since such biases may perpetuate societal inequalities by limiting access to well-paid jobs for people from lower social classes and hindering their upward social mobility.

4 - Hierarchy in Design Reviews: How Power and Status Influence Design Decisions

Yakira Mirabito, University of California, Berkeley, Berkeley, CA, United States, Kosa Goucher-Lambert

Engineers are systematic problem solvers whose decisions are not immune to the influence of biases and organizational dynamics. Authority bias is the tendency to overvalue opinions from the top of the hierarchy and undervalue those from the bottom. Hierarchy—defined here as power and status—has been observed in design education through instructor presence and feedback. Power refers to an individual's capacity to modify others' states via resources, while status denotes the respect others afford an individual. This study investigates the subtle dynamics of hierarchy within the context of a design review, an integral part of the engineering design course. Data from five industry design reviews across three engineering firms serve as the primary source, supplemented by semi-structured interviews and demographic surveys. Thematic analysis was employed to analyze the transcripts from meetings and interviews, explaining how verbal interactions in design reviews convey hierarchy. The act4teams coding schema was used to help map power and status to previously defined interactions. Meeting roles and feedback are two mechanisms through which power and status were wielded and influenced design decisions. Presenting designers, facilitators, and attendees have the potential to drive design decisions through the content under review and the feedback provided. The most influential feedback in guiding design decisions was linked to high-power or high-status individuals. This research deepens our understanding of how hierarchy impacts engineering design processes and proposes organizational changes to enhance equity and inclusion in design decisions.

SB02

Summit - 321

QSR Best Student Paper Competition

Award Session

Quality, Statistics and Reliability

Chair: Bing Si, State University of New York at Binghamton, Binghamton, NY, United States

Co-Chair: Chen Zhang, Tsinghua University, Beijing, N/A, China, People's Republic of

Co-Chair: Cesar Ruiz, University of Oklahoma, SILVER SPRING, MD, United States

1 - Transfer Learning of Stochastic Kriging for Individualized Prediction

Jinwei Yao, University of Iowa, Iowa City, IA, United States, Jianguo Wu, Yongxiang Li, Chao Wang

coming soon

2 - Explainable Parameter Calibration via Importance-Driven Sliced Sequential Design

Cheoljoon Jeong, University of Michigan, Ann Arbor, MI, United States, Eunshin Byon

coming soon

3 - Consistent Discovery of Dynamical Systems Using Stochastic Inverse Modeling

Ridwan Olabiyi, Arizona State university, Tempe, AZ, United States, Han Hu, Ashif Iquebal

coming soon

4 - Zero-shot Standard Operation Generation based on Pretrained Large Vision-language Model

Yifin Li, Tsinghua University, Beijing, China, People's Republic of, Xiaowei Yue, Hui Chen, Li Zheng

coming soon

SB03

Summit - 322

Modeling, Learning, and Leveraging Heterogeneous Information in Smart and Connected Systems

Invited Session

Quality, Statistics and Reliability

Chair: Chao Wang, University of Iowa, Iowa City, IA, 52242, United States

1 - Transfer Learning of Stochastic Kriging for Individualized Prediction

Jinwei Yao, University of Iowa, Iowa City, IA, United States, Chao Wang, Jianguo Wu, Yongxiang Li

Stochastic Kriging (SK) is a generalized variant of Gaussian process regression, and it is developed for dealing with non-i.i.d. noise in functional responses. Although SK has achieved substantial success in various engineering applications, its intrinsic modeling strategy by focusing on the sample mean limits its flexibility and capability of predicting individual functional samples. Moreover, the performance of SK can be impaired under scarce data scenarios, which are commonly encountered in engineering applications, especially for start-up or just deployed systems. In this paper, we propose a novel transfer learning framework to address the challenges of individualization and data scarcity in traditional SK. The proposed framework features a within-process model to facilitate individualized prediction and a between-process model to leverage information from related processes for resolving the issue of data scarcity. The within- and between-process models are integrated through a tailored convolution process, which quantifies interactions within and between processes using a specially designed covariance matrix and corresponding kernel parameters. Statistical properties are investigated on the parameter estimation of the proposed framework, which provide theoretical guarantees for the performance of transfer learning. The proposed method is compared with benchmark methods through various numerical and real case studies, and the results demonstrate the superiority of the proposed method in dealing with individualized prediction of functional responses, especially when limited data are available in the process of interest.

2 - Federated Multiple Tensor-on-Tensor Regression (Fedmtot) for Multimodal Data Under Data-Sharing Constraints

Zihan Zhang, ISyE Georgia Tech, Atlanta, GA, United States, Shancong Mou, Mostafa Reisi, Massimo Pacella, Jianjun Shi

In recent years, diversified measurements reflect the system dynamics from a more comprehensive perspective in system modeling and analysis, such as scalars, waveform signals, images, and structured point clouds. To handle such multimodal structured high-dimensional (SHD) data, combining a large amount of data from multiple sites is necessary (i) to reduce the inherent population bias from a single site and (ii) to increase the model accuracy. However, impeded by data management policies and storage costs, data could not be easily shared or directly exchanged among different sites. Instead of simplifying or facilitating the data query process, we propose a federated multiple tensor-on-tensor regression (FedMTOT) framework to train the individual system model locally using (i) its own data and (ii) data features (not data itself) from other sites. Specifically, federated computation is executed based on alternating direction method of multipliers (ADMM) to satisfy data-sharing requirements, while the individual model at each site can still benefit from feature knowledge from other sites to improve its own model accuracy. Finally, two simulations and two case studies validate the superiority of the proposed FedMTOT framework.

3 - Fourier Neural Operators for Fast Simulations in Additive Manufacturing

Raghav Gnanasambandam, Virginia Tech, Blacksburg, VA, United States, Zhenyu Kong

Metal additive manufacturing (AM) is a fast-growing technique that solves some unique problems in manufacturing. One of the ubiquitous issues with AM users is the unknown optimal process parameters that can provide high-quality parts. Physics-based simulations, repeated over several process parameter settings, are typically used for finding these optimal parameters. This work proposes to use the Fourier Neural Operator (FNO)-based predictions to optimize the process parameters in metal AM. Compared to standard deep-learning methods, FNO trains with limited physics-based simulations by learning the entire class of Partial Differential Equations (PDEs). Once trained, FNO instantaneously predicts the high-dimensional spatial-temporal output of physics-based simulations, typically the thermal distributions, with the process parameters as input. The accuracy of FNO is significantly better than that of existing methods for predicting simulation outputs. The higher accuracy is due to the lack of a principled way to handle physics-based functional data in standard deep-learning techniques. The case studies show that the proposed method can significantly reduce the simulation costs for optimizing the process parameters.

4 - Privacy-Aware System Failure Prediction Through Joint Longitudinal-Survival Modeling

Yuxin Wen, Chapman University, Orange, CA, United States, Miles Milosevich, Junde Chen

As manufacturing becomes increasingly data-driven, data privacy will become increasingly important for protecting sensitive data and ensuring the security and privacy of manufacturing operations. In this paper, we propose a novel data-driven joint longitudinal-survival modeling framework, which combines statistical survival models and federated learning methodologies to jointly model and extract degradation features, to predict the potential failure just-in-time while keeping the signals private.

5 - On Model Compression for Neural Networks: Framework, Algorithm, and Convergence Guarantee

Chenyang Li, New Jersey Institute of Technology, Newark, NJ, United States, Bo Shen

Model compression is a crucial part of deploying neural networks (NNs), especially when the memory and storage of computing devices are limited in many applications. This paper focuses on two model compression techniques: low-rank approximation and weight pruning in neural networks. However, training NNs with low-rank approximation and weight pruning always suffers significant accuracy loss and convergence issues. In this paper, a holistic framework is proposed for model compression from a novel perspective of nonconvex optimization by designing an appropriate objective function. Then, we introduce NN-BCD, a block coordinate descent (BCD) algorithm to solve the nonconvex optimization. One advantage of our algorithm is that an efficient iteration scheme can be derived with closed-form, which is gradient-free. Therefore, our algorithm will not suffer from vanishing/exploding gradient problems. Furthermore, with the Kurdyka-Łojasiewicz (KŁ) property of our objective function, we show that our algorithm globally converges to a critical point at the rate of $O(1/k)$, where k denotes the number of iterations. Lastly, extensive experiments with tensor train decomposition and weight pruning demonstrate the efficiency and superior performance of the proposed framework.

SB04

Summit - 323

Spatial-Temporal Event Data Analytic

Invited Session

Quality, Statistics and Reliability

Chair: Fenglian Pan, University of Arizona, Tucson, AZ, United States

Co-Chair: Jian Liu, University of Arizona, Tucson, AZ, United States

1 - Nondestructive Fatigue Life Prediction for Additively Manufactured Parts through a Multimodal Transfer Learning Framework

Jia Liu, Auburn University, Auburn, AL, United States, Shehzaib Irfan

Understanding the fatigue behavior and accurately predicting the fatigue life of laser powder bed fusion (L-PBF) parts remain a pressing challenge due to complex failure mechanisms, time-consuming tests, and limited fatigue data. This study proposes a physics-informed data-driven framework, namely, a multimodal transfer learning (MMTL) framework, to understand process-defect-fatigue relationships in L-PBF by integrating various modalities of fatigue performance, including process parameters, XCT-inspected defects, and fatigue test conditions. It aims to leverage a pre-trained model with abundant process and defect data in the source task to predict fatigue life nondestructively with limited fatigue test data in the target task. MMTL employs a hierarchical graph convolutional network (HGCN) to classify defects in the source task. The synergies learned from HGCN are then transferred
to fatigue life modeling in neural network layers. MMTL validation through numerical simulations and real-case studies demonstrates its effectiveness in fatigue life prediction of L-PBF parts.

2 - Block No-U-Turn Samplers for Spatially Correlated Accelerated Failure Time Models

Jie Min, Virginia Tech, Blacksburg, VA, United States, Yili Hong

The spatially correlated accelerated failure time (SC-AFT) model has received increasing attention over the years, due to its ability of utilizing important spatial information provided in time-to-event data. Moreover, Markov Chain Monte Carlo (MCMC) is considered a convenient method for the estimation and inference of SC-AFT models. Among various MCMC samplers, the No-U-Turn Sampler (NUTS) is widely used in practice because of its advantage in generating samples with high effective sample sizes. However, the computing time of NUTS can be long for sampling from SC-AFT models, partially because of the large number of parameters in the model. To solve the problem, we propose Block No-U-Turn Samplers (BNUTS) that are able to generate samples with reasonable auto-correlations within an acceptable amount of time for SC-AFT models. Two blocking schemes are proposed based on the SC-AFT model structure and the derivative computing method used in the sampler. Advantages of BNUTS over NUTS in sampling from SC-AFT models are demonstrated using extensive simulation examples. Important factors that affect the computing time and effective sample sizes of BNUTS and NUTS are further investigated. The proposed BNUTS is applied to analyze a GPU failure time data from the Titan supercomputer, showing the capability of BNUTS in solving real-world research problems.

3 - Detecting Changes in High Dimensional Model Relationship Profiles

Alina Gorbunova, Georgia Institute of Technology, East Brunswick, NJ, United States, Jianjun Shi, Kamran Paynabar

Advancements in sensor and data collecting technology have resulted in data sets that are both high dimensional and heterogeneous, with forms such as scalars, waveform signals, images, videos, and 3D point clouds. These advancements have created the need to construct and monitor statistical models that are able to model the relationship between heterogeneous and high dimensional inputs and a high dimensional output. This project (1) discusses how to model the relationship between such inputs and output using Multiple Tensor-on-Tensor Regression (MTOT) and (2) presents different approaches for how to monitor this model relationship to detect potential changes over time. Monitoring the model relationship can be done either by monitoring the core tensor of the MTOT model and/or by monitoring the residuals of new data samples. This methodology is validated through several case studies including monitoring overlay error from semiconductor wafer shape data and.

4 - Modeling Acute Opioid Event Recurrence with a Covariate-Adjusted Triggering Point Process

Fenglian Pan, University of Arizona, Tucson, AZ, United States, You Zhou, You Zhou, Carolina Vivas-Valencia, Nan Kong, Carol Ott, Mohammad S. Jalali, Jian Liu

Substance use disorder, particularly opioid-related, is a serious public health challenge in the U.S. Accurately predicting events related to opioid use disorder (e.g., overdose, intoxication) and stratifying the risk of having an event are critical to effective prevention and treatment for people with opioid use disorder. Despite a large body of literature investigating various risk factors for the prediction, the existing research to date has not explicitly investigated and quantitatively modeled how an individual's past events related to opioid use disorder affect future occurrences. In this paper, we propose a covariate-adjusted triggering point process to model the occurrence of recurrent acute opioid events that includes opioid use disorder diagnosis with withdrawal, intoxication, and poisoning, wherein an occurred acute opioid event is considered to trigger subsequent acute opioid events. Further, the proposed method is capable of incorporating the influence of identified demographic factors (e.g., age, sex, and race) in acute opioid event prediction. The prediction performance is assessed by a real-world case study. Compared with commonly used prediction models, the proposed method yields the lowest prediction error on 15-, 30-, 60-, and 90-day ahead predictions. In addition, our method can explicitly quantify the expected number of acute opioid events that are due to event triggering. Our results show the statistical significance of considering the triggering mechanism for recurrent acute opioid events prediction.

SB05

Summit - 324

INFORMS JFIG Best Paper Competition I

Award Session

Junior Faculty Interest Group

Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Disclosing Low Product Availability: An Online Retailer's Strategy for Mitigating Stockout Risk

Dmitry Mitrofanov, Boston College, Chestnut Hill, MA, United States

Ensuring product availability and the successful fulfillment of orders are key priorities for any company operating in the retail industry. In this paper, we investigate how sharing information regarding the low availability of certain items can influence customers' purchase decisions, both preventing stockouts and mitigating the negative effects of stockouts with respect to customer long-term behavior. It is hard to predict the net impact of sharing item availability information on business metrics ex-ante because there are multiple effects that might act in opposite directions. The information-sharing policy could lead customers to not purchase items that are low in availability because those customers are averse to stockouts. Contrariwise, this approach might increase the popularity of low-availability items because consumers' cognitive bias leads them to place a higher value on items that are scarce. In a field experiment, we exogenously share low item availability information with a random subset taken for a sample of more than 840K customers using Instacart. We find evidence that customers are 25% less likely to purchase low-availability items when item availability information is disclosed. In addition, we show that low product availability disclosure positively affects the platform's fundamentals in the context of our field experiment both in the short and long run. More specifically, our results indicate that this innovative and cost-free approach leads to a 5.33% increase in revenue per customer and a 4.9% increase in order frequency over the long term.

2 - Strategic Network Inspection with Location-Specific Detection Capabilities

Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States, Bastián Bahamondes

We consider a two-person network inspection game, in which a defender positions a limited number of detectors to detect multiple attacks caused by an attacker. We assume that detection is imperfect, and each detector location is associated with a probability of detecting attacks within its set of monitored network components. The objective of the defender (resp. attacker) is to minimize (resp. maximize) the expected number of undetected attacks. To compute Nash Equilibria (NE) for this large-scale zero-sum game, we formulate a linear program with a small number of constraints, which we solve via column generation. We provide an exact mixed-integer program for the pricing problem, which entails computing a defender's pure best response, and leverage its supermodular structure to derive two efficient approaches to obtain approximate NE with theoretical guarantees: A column generation and a multiplicative weights update (MWU) algorithm with approximate best responses. To address the computational challenges posed by combinatorial attacker strategies, each iteration of our MWU algorithm requires computing a projection under the unnormalized relative entropy. We provide a closed-form solution and a linear-time algorithm for the projection problem. Our computational results in real-world gas distribution networks illustrate the performance and scalability of our solution approaches.

3 - Learning to Defer in Content Moderation: The Human-AI Interplay

Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States

Ensuring successful content moderation is vital for a healthy online social platform where it is necessary to responsively remove harmful posts without jeopardizing non-harmful content. Due to the high-volume nature of online posts, human-only moderation is operationally challenging, and platforms often employ a human-AI collaboration approach. A typical machine-learning heuristic estimates the expected harmfulness of incoming posts and uses fixed thresholds to decide whether to remove the post (classification decision) and whether to send it for human review (admission decision). This can be inefficient as it disregards the uncertainty in the machine-learning estimation, the time-varying element of human review capacity and post arrivals, and the selective sampling in the dataset (humans only review posts filtered by the admission algorithm).

We introduce a model to capture the human-AI interplay in content moderation. The algorithm observes contextual information for incoming posts, makes classification and admission decisions, and schedules posts for human review. Non-admitted posts do not receive reviews (selective sampling) and admitted posts receive human reviews on their harmfulness. These reviews help educate the machine-learning algorithms but are delayed due to congestion in the human review system.
We propose a near-optimal learning algorithm that carefully balances the classification loss from a selectively sampled dataset, the idiosyncratic loss of non-reviewed posts, and the delay loss of having congestion in the human review system. To the best of our knowledge, this is the first result for online learning in contextual queueing systems and hence our analytical framework may be of independent interest.

SB06

Summit - 325

Facility/Hub Location

Invited Session

Location Analysis

Chair: Gita Taherkhani, Loyola University Chicago, Chicago, IL, United States

1 - Time-Dependent Decisions in Hub Location and Routing

Francisco Saldanha-da-Gama, Sheffield University Management School, Sheffield, United Kingdom, Afaf Alloulal, Raca Todosijevic

This paper investigates the relevance of time-dependent decisions in hub-location routing problems. A multi-period planning horizon is considered during which the system is to be phased in. In addition to hub location, hub network decisions are also considered. The origin-destination flows are assumed to be time-dependent as well as the costs underlying the problem which include, setup costs for hubs and hub edges and variable operational costs at the hubs. An optimization model is discussed for the problem. For small instances, it can be solved up to proven optimality using a general-purpose solver. For larger instances, a four-phase matheuristic combining principles of relax-and-fix, variable neighborhood descent and local branching schemes is proposed. Two variants of the algorithm are discussed. The model and proposed procedure are tested using data generated by extending existing hub location instances to our problem. The results are analyzed. The major conclusion to draw is that by capturing time in the decision-making process, one may find solutions that better hedge against parameter changes throughout time. Furthermore, the overall procedure presented in this paper is quite general in the sense that it can be easily adapted to other multi-period decision making problems and different objective functions.

2 - Hub Network Design for Platooning

Elif Zeynep Serper, TED University, Ankara, Turkey, Sibel Alumur Alev

In transportation, platooning is a method where multiple vehicles, typically trucks, travel closely together in a convoy led by one vehicle with others following at close distances. This formation is maintained by advanced automated driving systems. Platooning is primarily utilized to enhance both fuel and driver efficiency. Through closely trailing the lead vehicle, subsequent vehicles experience reduced aerodynamic drag, resulting in decreased fuel consumption. Furthermore, as trucks form platoons, driver costs are significantly reduced since drivers can rest in the truck, enabling continuous travel without mandated breaks. The advantages of platooning include cost reduction due to fuel savings and increased driver efficiency, reduced carbon emissions contributing to environmental sustainability, increased safety, driver comfort, and improved traffic flow by reducing the space between vehicles. In platooning applications, hubs serve as roadside locations where trucks can gather, wait, and form or deform platoons. In this study, we focus on the design of hub networks for truck platooning applications in truckload transportation. We develop a mixed-integer programming model to optimize the locations of hubs considering the trade-off between platooning and direct shipment in terms of both service time and cost. The model determines whether a truckload is to be shipped using platooning, the optimal locations of platoon formation, and deformation hubs while adhering to predetermined service time restrictions and platoon capacities. The model is tested on real-life data from U.S. trucking operations.

3 - Service Network Design and Hub Location

Gita Taherkhani, Loyola University, Chicago, IL, United States, Hao Li, Sibel Alumur Aleb, Mike Hewitt

To enhance cost-effectiveness, it is essential that shipments transported by carriers are consolidated within a hub network. Additionally, the movement of these shipments, executed by capacitated trucks, must account for spatial and temporal decisions typically considered in service network design. This paper introduces a profit-maximizing model that integrates hub location and service network design for less-than-truckload (LTL) carriers. The model strategically selects hub locations and makes tactical decisions regarding truck dispatch times, streamlining the process without needing to model every time point explicitly. A deterministic model and its Benders reformulation are presented, enhanced with acceleration techniques. The model is extended to handle variability in shipment volumes through a two-stage stochastic program. To solve this model, a Monte-Carlo simulation-based algorithm integrated with a sample average approximation scheme and a proposed Benders decomposition reformulation is developed. Computational results demonstrate that this formulation can solve instances involving up to 20 nodes over a three-day planning period. Its performance is compared with that of commercial solvers and other algorithms.

4 - Shipment Consolidation in Multi-Period Hub-Arc Location Problems with Stochastic Demand

Jim Bookbinder, University of Waterloo, Toronto, ON, Canada, Khaled Shah, Sibel Alumur Alev

We include inventory holding decisions in hub arc location problems with probabilistic origin-destination demands. A heterogeneous fleet, with varied capacities and costs are used on the hub arcs. Vehicles may be used with different ‘frequencies’, meaning the number of vehicles to be operated on a given arc may vary during different days (periods) of the time horizon. We formulate a mixed-integer model that minimizes the total costs of hub arc establishment, transportation plus inventory holdover, given the demands and a choice of frequencies for the hub arcs. We adopt the sample average approximation approach to deal with demand randomness. The performance of our model and algorithm are evaluated by numerical experiments.

SB07

Summit - 327

Social Responsibility in Operations

Flash Session

Contributed

Chair: Chenwei Jin, N/A

1 - Advancing Carbon Mitigation via U.S. Municipal Green Bond Issuance: Insights from Causal Machine Learning

Dan Li, university of michigan, Ann Arbor, MI, United States, Peter Adriaens

The green bond market provides a vital channel for directing financial resources toward climate-related initiatives that support the Sustainable Development Goals. This study aims to evaluate the influence of the US municipal green bond market on carbon emissions spanning from 2009 to 2019, taking into account the confounding effects of local socio-economic conditions. By employing causal machine learning models, we discerned the effects of increased green bond issuance volume on emission reduction at the county level. Our findings reveal notable heterogeneity across regions, with a maximum impact on emissions reduction two years post-bond issuance. Within the subgroups analyzed in our study, counties characterized by long commute times appear to benefit most significantly. Hence, urbanization and city expansion may result in a notably larger impact from green bond issuances to mitigate CO_2 emissions. Importantly, our findings highlight the effectiveness of certified green bonds, which undergo environmental outcome verification by assurance companies, in catalyzing efforts to reduce emissions. While this verification process undoubtedly incurs costs for the external review, it also enhances the credibility of their environmental commitment. These novel insights can inform the benefit of green bond issuance and strategic allocation of resources for emissions mitigation to regions that most likely to derive benefits from targeted environmental measures and urban planning initiatives.

2 - Evolution of gender pay gap in the 21st century

Jeena Ahuja, Alcuin School, Dallas, TX, United States

The gender wage gap has been studied for years, but it remains an area of active and innovative research. This research strives to better understand the worldwide historical trends in gender wage gap \over the course of two decades -- from 2000 to 2020. In particular, we compare and contrast ten different regions -- Oceania, two in Americas (North, South), four regions in Europe (Eastern, Western, Northern, Southern), Asia, and Middle East. We also benchmark against the average of Organization for Economic Co-operation and Development (OECD) countries. We use the publicly available OECD dataset to conduct our analysis. We use multiple regression models created in *RStudio* and F-tests to conduct our analysis. We find that the gender wage gap follows a quadratic trend in each region. Importantly, the gender wage gap decreased over the course of two decades in all but Oceania, Northern Europe, and South America. Additionally, we showed (using t-tests) how significant the changes in the gender wage gap are between two consecutive years. Specifically, we find that the most significant changes in the evolution of the gender wage gap occurred in the second decade, specifically in the years 2010-2011, 2014-2015, and 2018-2019. In summary, we analyze the evolution of gender wage gap in the first two decades of the 21st century and the associated regional and temporal variation. Our findings highlight the areas that need more attention from policymakers.

3 - Analysis of Corporate Environmental, Social, and Governance Performance

Xinyi Lai, Westwood High School, Austin, TX, United States

Environmental and social responsibilities play pivotal roles in fostering the growth and advancement of our society. With increasing awareness of the pressing need for sustainable practices, there is a widespread belief that companies emphasizing environmental, social, and governance (ESG) factors can not only enhance their reputations but also attract and retain top-tier talent, contributing to a more sustainable and equitable world. However, despite the recognized importance of ESG initiatives, the extent to which companies are effectively addressing these responsibilities remains uncertain. To shed light on this issue, this study collects reported ESG scores from publicly traded companies and evaluates their overall performance in fulfilling their ESG obligations. Additionally, this study explores factors that may influence companies' ESG scores and examines the connections between companies' ESG performance and their financial outcomes.

4 - The Importance of Investment in Electric Cars

Andrew Zheng, Boston Latin School, Boston, MA, United States

It is no secret that vehicle exhaust is extremely harmful for us and the environment. Electric cars can help mitigate this problem as they have low to no tailpipe emissions whatsoever. Pollutants from vehicle exhaust cause at least 100,000 premature deaths per year in the United States alone. Fortunately, the world is aware of this and many countries are rushing to transition into EVs. For example, in Norway, over 80% of new car sales are EVs. Other European countries are catching up, and if the car market follows current predictions, by 2030 Europe could save emissions equivalent to planting one billion trees. If all of the cars on the road become electric, global emissions could be cut by 20%. National policies and incentives can help push EV sales in other parts of the globe. These policies and incentives include subsidies in the form of tax credits to consumers, as this now helps alleviate the higher upfront costs of purchasing EVs. Currently, the U.S. government offers tax credits ranging from $2,500 to $7,500. These tax credits are responsible for about 30 percent of carbon dioxide emissions reductions from the current amount of EVs in the U.S. Mandates could also decarbonize the car industry, as seen in California's Zero Emission Vehicle(ZEV) Sales Requirement program, which forces automakers to gradually make all of their cars electric by 2035. These ideas would definitely reduce our emissions and help save our planet from the irreversible effects of climate change.

5 - The impact of built environment and private car accessibility on transit-oriented-development in urban sub-centres: A case study of Shanghai's sub-centre metro stations

Chenwei Jin, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

A number of studies have suggested that TOD (transit-oriented development) in sub-centres with transit stations is an approach to increase foot traffic and the overall vitality of the area. However, the current development of sub-centres mostly follows the existing well-established TOD model in CAZs (central activity zones), which may not be applicable to sub-centre areas. Additionally, most TOD studies do not deliberately distinguish or emphasise the differences between CAZs and sub-centres. Therefore, this study aims to examine the influence of various factors in the built-up areas of Shanghai's urban sub-centres on the degree of TOD around the stations and to assess the level of development at existing transit stations. The four main factors considered are accessibility to private cars, land use, inner connectivity, and functionality. A regression model was introduced to identify and validate the factors that significantly influence the usage of transit stations, as reflected by the daily passenger flow at transit stations. The results show that almost all factors are associated with transit station usage, while land use diversity is not as significant as it is in urban centres.

SB08

Summit - 328

OR in Security & Defense I (Operations)

Invited Session

Military and Security

Chair: Paul Goethals, Department of Defense, Columbia, MD, United States

1 - Automated Vessel Selection and Combat Load Planning

William Kirschenman, North Carolina State University, Raleigh, NC, United States

In large-scale combat operations, the U.S. military must move and maneuver its forces through intratheater and intertheater modes of transportation. Threat forces exacerbate these demanding requirements through efforts to hinder the flow of friendly forces. Contested landing zones, whether they be fixed ports or beaches, are the starting point for a landing force’s ground combat operations. It is imperative that the landing force expeditiously off-loads in the prescribed order of priority to support the planned scheme of maneuver upon off-loading. We first discuss current and historical methods of load planning while emphasizing the need for more detailed and automated methods driven by the expected nature of future conflicts against near-peer threats. We then present a model that optimizes vessel selection, sequencing, and combat load configurations of a large military force by considering multiple levels of priority and group unity, which enables efficient off-loading into desired tactical formations for follow-on objectives.

2 - Resupply during Dispersed Operations

Mette Wagenvoort, Erasmus University Rotterdam, Rotterdam, Netherlands, Paul Bouman, Martijn van Ee, Kerry Malone

During a military operation, units require supplies such as water, food and fuel. Therefore, during the operation, the units need to be resupplied to ensure they are never without supplies. In a dispersed operation, small units operate at geographically disperse locations, making the resupply task more challenging.

There are different ways in which units can be resupplied. Namely, a large quantity of different supplies can be brought to a location from which units can collect them. Alternatively, smaller quantities of a commodity can be brought to one of the units. However, this is riskier as the supplies have to be brought closer to the area of operations. There is thus a trade-off between offering more tailored services at a higher risk for the resupply operation, and more generic services that require a higher time investment from the units itself.

When the region in which units operate is far from the depot, an intermediate depot can be placed from which units can be resupplied. The intermediate depot will have to be resupplied from the main depot, as it is not desirable to have large quantities of stock in the intermediate depot. However, this might require more vehicles to execute the resupply operation.

It is therefore of interest to model the resupply operation using different types of resupply methods to analyse in which situation one would prefer each of these methods.

3 - Resilient Taksing and Basing Model - an optimization of millitary resources placement in theater

Oluwaseun Ogunmodede, The MITRE Corporation, McLean, VA, United States

The Resilient Tasking and Basing Model (RTBM) is an attacker-defender model that finds the optimal beddown of tactical aircraft in the face of base attacks by a near-peer threat. Optimality here is a saddle-point where Blue can still accomplish its mission while minimizing its casualties. We discuss the structure of the antagonists. Blue has a multi-layered facility location structure: where to base receiver aircraft in order to service missions, which interacts with where to base tankers in order to service aerial refueling points that are servicing the receiver aircraft. Similarly, Red has a fixed inventory of missiles that must choose given launch points from which to attack Blue bases to kill Blue AC, maximizing casualties. We show how to transform this Min - Max Defender-Attacker optimization model into a tractable minimization optimization monolith.

4 - Adversarial Unsupervised Learning in Connected Defense Systems

Tahir Ekin, Texas State University, San Marcos, TX, United States

Traditional unsupervised learning methods typically assume clean and legitimate data streams. However, the presence of changing data, patterns and adversarial influences may impact their performance in dynamic environments. The intersection of dynamic data driven applications systems and adversarial risk analysis presents an opportunity for comprehending and effectively managing dynamic conditions with security and reliability. This manuscript proposes a decision-theoretic framework for clustering based outlier detection within the realm of dynamic unsupervised learning, addressing the challenges posed by adversarial attacks, uncertainty, and incomplete information. We demonstrate the efficacy of our framework within a joint all-domain command and control systems framework.

SB09

Summit - 329

Design and Analysis of Marketplaces

Invited Session

Auctions and Market Design

Chair: Chen Chen, New York University Shanghai, Shanghai, N/A, China, People's Republic of

1 - Budget Pacing in Repeated Auctions: Regret and Efficiency Without Convergence

Bar Light, Tel Aviv University, Tel Aviv, Israel, Brendan Lucier, Alex Slivkins

We study the aggregate welfare and individual regret guarantees of dynamic \emph{pacing algorithms} in the context of repeated auctions with budgets.
Such algorithms are commonly used as bidding agents in Internet advertising platforms.
We show that when agents simultaneously apply a natural form of gradient-based pacing, the liquid welfare obtained over the course of the learning dynamics is at least half the optimal expected liquid welfare obtainable by any allocation rule.
Crucially, this result holds \emph{without requiring convergence of the dynamics}, allowing us to circumvent known complexity-theoretic obstacles of finding equilibria.
This result is also robust to the correlation structure between agent valuations and holds for any \emph{core auction}, a broad class of auctions that includes first-price, second-price, and generalized second-price auctions. For individual guarantees, we further show such pacing algorithms enjoy \emph{dynamic regret} bounds for individual value maximization, with respect to the sequence of budget-pacing bids, for any auction satisfying a monotone bang-for-buck property.

2 - Trading Flexibility for Adoption: Dynamic Versus Static Walking in Ridesharing

Sebastien Martin, Kellogg School of Management, Northwestern University, Evanston, IL, United States

On-demand ridesharing aims to fulfill riders' transportation needs whenever and wherever they want. Although this service level appeals to riders, overall system efficiency can improve substantially if riders are willing to be flexible. Here, we explore riders' flexibility in space via walking to more accessible pickup locations.
Ridesharing platforms have traditionally implemented *dynamic walking* to optimize rider pickup locations and rider-driver assignment jointly. We propose an alternative that we call *static walking*, which presents a predetermined pickup location to the rider before optimizing rider-driver assignment. Although dynamic walking enables more efficient matching of riders and drivers, we hypothesize that riders prefer static walking because of the certainty of the pickup location before booking the ride. Using simulations on Lyft data, we show that static walking can capture up to 96% of the value of dynamic walking in congested urban networks at a fixed adoption rate. Furthermore, experimentation on Lyft's user interface suggests that providing riders with information on pickup location before an opt-in decision can increase walking adoption --- to the extent that static walking may outperform dynamic walking overall. More broadly, this study highlights the importance of carefully designing flexibility mechanisms on platforms: a little flexibility goes a long way, especially when flexibility presents a barrier to adoption.

3 - Incentivizing Resource Pooling

Chen Chen, New York University Shanghai, Shanghai, China, People's Republic of, Yilun Chen, Pengyu Qian

Resource pooling improves system efficiency drastically in large stochastic systems, but its effective implementation in decentralized systems remains relatively underexplored. This paper studies how to incentivize resource pooling when agents are self-interested, and their states are private information. Our primary motivation is applications in the design of decentralized computing markets, among others. We study a standard multi-server queueing model in which each server is associated with an M/M/1 queue and aims to minimize its time-average job holding and processing costs. We design a simple token-based mechanism where servers can earn tokens by offering help and spend tokens to request help from other servers, all in their self-interest. The mechanism induces a complex game among servers. We employ the fluid mean-field equilibrium (FMFE) concept to analyze the system, combining mean-field approximation with fluid relaxation. This framework enables us to derive a closed-form characterization of servers' FMFE strategies. We show that these FMFE strategies approximate well the servers' rational behavior. We leverage this framework to optimize the design of the mechanism and present our main results: As the number of servers increases, the proposed mechanism incentivizes \emph{complete} resource pooling---that is, the system dynamics and performance under our mechanism match those under centralized control. We also extend our mechanism to settings with heterogeneous servers and show that our mechanism obtains near-optimal performance.

4 - The Welfare Effects of Selling Leads in a Two-Sided Marketplace

Peng Shi, USC Marshall School of Business, Los Angeles, CA, United States

Digital platforms that help customers find suitable service providers often monetize by selling customer leads to interested service providers. Examples of such platforms include Bark, Google Local Services, HomeAdvisor, Modernize, Porch, and Thumbtack. I analyze this platform design using a game-theoretic model and obtain insights on how the pricing of leads affects customer and provider welfare. When the platform raises the fee per lead, providers who buy these leads quote more competitive prices to customers to improve their chances of converting these leads. This benefits the customers whose leads are bought, but more customers may be unmatched because no one buys their lead. For providers, raising the fee per lead allows them to more easily target this customer segment, but those who successfully target receive a lower surplus. For maximizing social welfare, the simple policy of charging a market-clearing fee-per-lead for every customer segment guarantees at least 1/(e-1)≈58.19% of the first-best welfare, and at least 79.15% of the welfare under the optimal fees. Higher welfare can be achieved by paying providers an additional subsidy upon each service, so that experienced providers with cheaper sources of leads do not exit the platform. Understanding the nuances of these welfare effects can help platforms better grow both sides of the marketplace while maintaining an adequate revenue stream.

SB10

Summit - 330

Frontiers of Platform Design

Invited Session

Auctions and Market Design

Chair: Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States

Co-Chair: Mika Sumida, University of Southern California, Los Angeles, CA, United States

1 - Two-Sided Feature-Based Choice Modeling via Optimal Transport

Yijie Zheng, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu

Conventional choice models typically concentrate on choice decisions from one side, such as consumers' product choices, which makes it hard to handle scenarios where both sides simultaneously make choices. Moreover, these one-sided models generally use feature information solely from the chosen side, failing to accommodate the diversity of both sides involved. Inspired by optimal transport theory, this paper introduces a two-sided feature-based choice modeling approach to tackle the two challenges. Within the two-sided choice model, the interactions between consumers and products are conceptualized as a mutual choice process, akin to a matching mechanism. Consumer attributes are projected onto the product feature space through a random preference matrix, facilitating the matching of consumers with their nearest products. By allowing an arbitrary random preference matrix, our model demonstrates extensive generality, encompassing commonly used choice models. In the context of multi-product pricing, we demonstrate under mild conditions that optimal prices maintain a constant markup over their unit costs and that the profit function exhibits concavity with respect to market shares, enabling efficient price optimization across various scenarios. For assortment planning, we develop an efficient solution method using the shortest path approach, which solves problems in polynomial time. Additionally, we use a nonparametric estimation method based on the conditional gradient descent algorithm to learn the random preference matrix and demonstrate its effectiveness both theoretically and empirically.

2 - Platform Design for the First Mile of Commodity Supply Chains

Sergio Camelo, Stanford University, Stanford, CA, United States, Dan Iancu, Joann de Zegher

We propose a data-driven platform that provides traceability to the first mile of agricultural supply chains by coordinating the transactions of farmers and intermediaries. We model unique aspects of the supply chain, including pre-existing informal relationships between farmers and intermediaries, and we develop algorithms to solve real-world instances. We test the results on data from the palm oil supply chain and show the platform's potential to reduce costs and increase farmers' welfare.

3 - Hicksian Unit-Demand Pricing

Michael Hamilton, University of Pittsburgh, Pittsburgh, PA, United States, Jourdain Lamperski, Kasra Tari

We study single price approximations for the Hicksian unit-demand pricing problem. In the well-studied Bayesian unit-demand pricing problem, two celebrated lines of work show that (i) when valuations are IID, an optimally chosen single price is a two-approximation of the optimal pricing, and (ii) when valuations are independent but not identical and satisfy an MHR condition, a single price is a $\approx$ 2.62 approximation of the optimal pricing. We study the equivalent problem when customers choose items in a Hicksian (sometimes referred to as min-buying) fashion, purchasing the minimum priced item that guarantees a certain level of utility. We find that under Hicksian demand and IID valuations, a single price is always optimal, implying that the paradoxical advantage of non-symmetric prices vanishes in more conservative demand models. We also give improved single price approximations in the case of independent but non-identical valuations, and new results for the case of identical but dependent exchangeable valuations.

4 - Optimal world design in video games

Lifei Sheng, University of Houston Clear Lake, Houston, TX, United States, Yifu Li, Christopher Ryan

Spending time in virtual spaces is a growing part of human experience. We study the design of virtual spaces in a video game context, with an emphasis on understanding how people spend more or less time enjoying these spaces. We model virtual world design as a graph design problem. We show that, in certain cases of virtual world design, the optimal world maps have a ``side-quest'' tree structure that is amendable to an efficient inductive construction using a polynomial-time algorithm. We also show numerically that side-quest trees are ``near'' optimal for more general cases.

5 - Tracking Truth with Liquid Democracy

Daniel Halpern, Harvard University, Cambridge, MA, United States

The dynamics of random transitive delegations on a graph are of particular interest when viewed through the lens of an emerging voting paradigm, *liquid democracy*. This paradigm allows voters to choose between directly voting and transitively delegating their votes to other voters, so that those selected cast a vote weighted by the number of delegations they received. In the epistemic setting, where voters decide on a binary issue for which there is a ground truth, previous work showed that a few voters may amass such a large amount of influence that liquid democracy is less likely to identify the ground truth than direct voting. We quantify the amount of permissible concentration of power and examine more realistic delegation models, showing they behave well by ensuring that (with high probability) there is a permissible limit on the maximum number of delegations received. Our theoretical results demonstrate that the delegation process is similar to well-known processes on random graphs that are sufficiently bounded for our purposes. Along the way, we prove new bounds on the size of the largest component in an infinite P\'olya urn process, which may be of independent interest. In addition, we empirically validate the theoretical results, running six experiments (for a total of $N=168$ participants, 62 delegation graphs and over $11k$ votes collected). We find that empirical delegation behaviors meet the conditions for our positive theoretical guarantees. Overall, our work alleviates concerns raised about liquid democracy and bolsters the case for the applicability of this emerging paradigm.

SB11

Summit - 331

New Applications of Network Optimization

Invited Session

Telecommunications and Network Analytics

Chair: Rui Zhang, University of Colorado Boulder, Boulder, CO, 80309, United States

Co-Chair: Yunlong Wang, Shanghai Jiao Tong University. University of Colorado Boulder (Visiting), 2300 Arapahoe, Apt 2B 208 Netwon Court , Shanghai, 200030

1 - New Formulations and Valid Inequalities for the Least Cost Influence Problem on Social Networks

Yunlong Wang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, TaoTao He, S. Raghavan, Ying Rong, Rui Zhang

This work studies the least cost influence problem where the decision maker wants to promote a new product over a given social network, and the goal is to have every individual adopt this product with the least cost. In the given network, each individual receives influence from their neighbors who have adopted the products, and they also may influence others once they become adopted. A threshold is applied to the activation condition. We developed several IP formulations and designed valid inequalities to enhance them. Theoretical results are given, and comprehensive numerical experiments are implemented to certify the promise of the proposed methods. In addition, we conducted a thorough study regarding the optimal promotion strategy on different network structures and gave some managerial insights into this problem.

2 - Compact Mip Formulations for the Minimum Biclique Cover Problem

Hamidreza Validi, Texas Tech University, Lubbock, TX, United States, Bruno Burin, Bochuan Lyu, Illya Hicks

Given a simple graph G = (V, E) with vertex set V and edge set E, the minimum biclique cover problem seeks to cover all edges of the graph with a minimum number of bicliques (i.e., complete bipartite subgraphs). This is an NP-hard combinatorial optimization problem with applications in disjunctive programming, fixing decision variables in mixed integer programming (MIP), and combinatorial geometry. This talk proposes two compact MIP formulations for solving the minimum biclique cover problem on general graphs: (i) a natural formulation in the edge space and (ii) an extended formulation in the edge and vertex spaces. Despite the exponential-size MIP formulation of Cornaz and Fonlup (Discrete Mathematics, 2006), our natural formulation employs only a polysized number of their exponential "no-good" cuts, along with another set of polysize valid inequalities. Although our natural formulation has fewer decision variables, the proposed extended formulation performs better computationally. To accelerate the solving process of the extended formulation further, we propose (i) valid inequalities, (ii) fixing procedures, and (iii) a bounding procedure. Our computational experiments show the proposed framework's efficacy to improve the extended formulation's computational performance.

3 - Cactus Graph Formation Algorithms for Communication Networks with Decentralized Control

Alexander Nikolaev, University at Buffalo, Buffalo, NY, United States, Sangeeth Das, Robert Dell

A directed cactus graph, by definition, is such a graph where any directed edge is part of only one directed loop/cycle. On the theoretical front, this work contributes to the domain of network optimization by presenting computationally efficient heuristic solutions to the graph formation problem where the objective is to form a cactus graph with the minimum number of edges. The solutions obtained with our sub-optimal, low-computational-expense algorithms are shown to compare well to the optimal solutions on a number of cactus formation problem test instances. On the application front, we explain how the cactus graphs we are able to form can be used in communication network design applications, in particular to avoid double-counting in data fusion. Double-counting, or data incest, is a problem that arises in networks when members (sensors) of a decentralized network are tasked to make estimates based on shared signals. Our cactus graph formation-inspired decentralized algorithm is a self-organizing, low complexity approach that any placed sensors can adopt at the stage of establishing network-wide communication, thereby achieving connection structures enabling evidence exchange and aggregation with no double-counting.

4 - Lethal Cliques in Comorbidity Graphs

Parisa Vaghfi Mohebbi, Oklahoma State University, Stillwater, OK, United States, Yajun Lu, Zhuqi Miao, Baski Balasundaram, Pankush Kalgotra, Ramesh Sharda

Mortality rate refers to the overall likelihood of death within a specific population over a defined period. The knowledge of high mortality rate disease clusters can enable healthcare providers and patients to be proactive and develop tailored interventions that improve patient outcomes. In this talk, we discuss the problem of finding a small clique of comorbid diseases that corresponds to the highest mortality rate among a given patient population. We introduce two approaches to solve this problem: (i) an integer programming formulation that maximizes a single fractional objective subject to linear constraints, and (ii) an extension of the classical Bron-Kerbosch enumeration algorithm. We conduct a detailed computational study and report results from our experiments with both approaches on datasets derived from 10.6 million anonymous patient electronic health records.

5 - Routing and re-routing UAV's under Uncertain Disruptions

Jaime Yair Perez Tezoco, Oklahoma State University, Stillwater, OK, United States, Juan Borrero

We consider a UAV routing problem in which the visiting targets are subject to uncertain disruptions. The goal is to find a tour, limited in length, that maximizes the sum of collected information of selected locations. However, information cannot be collected if a target location is disrupted. Initially, the planner is unable to have information regarding the locations in which disruptions will occur, and this information is revealed once a disrupted target is visited. To address this problem, we propose a two-stage robust optimization model where the first-stage problem is to plan a tour under the uncertain environment, and the second stage problem is to optimize the re-routing recourse actions to be taken once the information regarding disrupted locations is revealed. We solve this problem by using a column-and-constraint generation algorithm and a decomposition branch-and-cut method. We present computational results considering both randomly generated networks and instances available in the literature for similar applications, which corroborates the effectiveness of the proposed model and the solution method.

SB12

Summit - 332

Logistics and Sustainable Transportation

Flash Session

Contributed

Chair: chang liu, Harbin Engineering University, Harbin, N/A, China, People's Republic of

1 - Advancing Decarbonization and Blockchain Integration in the Supply Chain: Strategic Decision-Making under Dual Carbon Regulatory Policies

chang liu, Harbin Engineering University, Harbin, China, People's Republic of

Excessive carbon emissions have led to an exacerbation of environmental issues. This study initially examines how supply chain members develop carbon reduction technology and select carbon reduction pathways under the dual carbon policies of carbon tax and cap-and-trade systems. Furthermore, it conducts a comparative analysis of the profits of supply chain members before and after the development of blockchain technology, quantifies the factors and investment thresholds that affect the maximization of corporate profits, and explores how to optimize the profits of supply chain members. Through the construction of game-theoretic models and numerical simulation analysis, it is found that the impact of carbon tax and cap-and-trade policy formulation on manufacturers' profits is related to the carbon emission values of the original product production and the selection of carbon reduction channels. The development of carbon reduction technology can significantly increase the profits of retailers. Under different market demands and carbon policies, manufacturers adopt different carbon reduction pathways. The choice of manufacturers' carbon reduction channels hinges on the balance between gains and cost investment. A comparison of investment costs and thresholds across different reduction modes guides the selection of the reduction mode. The introduction of blockchain technology has significantly improved the level of carbon reduction, and an optimal investment scheme for the blockchain platform has been identified. That is, when manufacturers and retailers jointly share the costs of blockchain technology within a threshold, a double Pareto improvement can be achieved.

2 - Optimizing Bulk logistics under uncertainty and compatibility constraints

Ashesh Kumar Sinha, West Virginia University, Morgantown, WV, United States

This study explores the optimization of routes for transporting bulk items like oil products, aerosols, and foodstuffs, which are typically loaded directly into tankers or trailers without conventional packaging. Unlike traditional logistics, Bulk logistics entails adhering to prior order restrictions, which are based on compatibility with recent orders (up to the last five), as well as potential requirements for specialized equipment or handling instructions, such as meticulously cleaned and prepared tanks. Furthermore, the research aims to incorporate two types of uncertainty into the model: (i) limited capacity at each wash location after delivery, and (ii) the degradation of order quality over time, resulting in diminished value upon delivery completion. By employing an integrated approach utilizing stochastic models and graph decomposition theories, we tackle the challenges associated with order compatibility and uncertainty. The first stage formulation draws upon graph decomposition theories to manage compatibility relations. We develop specialized graph decomposition-based theories that are proven to be optimal and derive tight lower bounds on the optimal solution to further enhance the decomposition-based approach. The second stage formulation analyzes uncertainty within the system and derives necessary optimality and feasibility constraints. Using numerical experiments, we study the impact of parameters such as number of trailers, wash locations, and various order compatibility scenarios on the optimal routing decisions.

3 - Fleet Scheduling for Distribution Planning of Refrigerated and Non-Refrigerated Loads

Ahmad Bassaleh, Wichita State University, Wichita, KS, United States, ekrem duman, Mehmet Yildirim

Logistical challenges pose a substantial financial burden for distributors worldwide, particularly in light of escalating oil prices and evolving customer demands. This paper addresses a Vehicle Routing Problem (VRP) encountered by a distributor in Turkey. The problem involves a multi-objective function that integrates road tariffs, specific customer constraints, refrigerated and non-refrigerated loads, and features observed in various VRP variants. Our study explores diverse solving approaches, including an exact method developed through a novel Mixed-Integer Linear Program, as well as non-exact solutions employing heuristic and metaheuristic algorithms. Notably, our findings highlight the efficacy of metaheuristic algorithms, especially within a smart two-phase solving strategy designed to tackle the problem's large scale: initially segmenting it into individual sub-problems and subsequently integrating and solving them comprehensively. The study concluded with a noteworthy 5.13% reduction in operational costs compared to the distributor's existing plan, indicating significant potential for enhanced profitability while optimizing the daily scheduling of the vehicle fleet.

4 - Biking the goods... but where?

Carla Tejada, University of Illinois in Chicago, Chicago, IL, United States

Biking in general, has seen an increase in several parts of the world. In addition, as more sustainable freight transportation modes are being developped, biking the goods has taken an interest from researchers and practitioners for the last portion of goods delivery. Additionally, several cities in the U.S. have seen an increase in the bike ridership that requires the development of biking infrastructure. However, the developped infrastructure has yet to be proved safe and it generally is not designed to accomodate for cargo-bikes.

This preliminary study has three specific goals. The first is to assess the perceived safety for cyclists in urban areas, with the already existing bycicle infrastructure. For this, an AI vide recognition model has been developped to recognize all the instances where cyclists pereived a nuissance. The second goal is to assess what is the existing connectivity of bike lanes. To accomplish this goal we have created a connectivity model that will help assess the existing infrastructure. To prove this model, data from the Chicago area was used. Finally, the last goal, is to create a qualitative assessment of whether or not the exisitng infrastructure is ready to accomodate cargo-bikes and what would be the requierements for this implementation in the future.

Initial results show the importance of identifying existing nuissane to improve the existing bike infrastructure, and has shown that a lot of work remains to be done to improve safety and capacity of bike infrastructure in cities.

5 - Bi-Objective Electric Vehicle Routing Problem with Queuing Systems at Charging Stations

Xin Zhang, Tongji University, Shanghai, China, People's Republic of, Jiantong Zhang

With the rise in battery capacities and the expansion of fast-charging infrastructure, electric vehicles (EVs) are increasingly suitable for long-distance travel. However, queues at charging stations may result in extended waiting times, underscoring the importance of efficient route planning. By integrating arrival rules, queuing rules, charging stations, and service times, a queuing system is established. This study aims to investigate the bi-objective Electric Vehicle Routing Problem (EVRP) while considering queuing systems at charging stations. A mixed-integer programming model is formulated to address the problem, and a non-dominated ranking genetic algorithm with an elite strategy is developed to solve it. Test results demonstrate that the model and algorithm proposed in this study efficiently identify the Pareto optimal solution set for the problem.

SB13

Summit - 333

Autonomous Logistics and Planning

Flash Session

Flash

Chair: Ali Tolooie, Mercer University, Macon, GA, United States

1 - A Branch-and-Price-and-Cut Type Algorithm for the Collaborative Truck&Ndash;Drone Routing and Scheduling Problem with Flexible Launch and Recovery Locations

SHROUQ GAMAL, KFUPM, Dharhan, Saudi Arabia, Naqeebuddin Mujahid Syed

This study introduces the Collaborative Truck–Drone Routing and Scheduling Problem with Flexible launch and recovery locations (CTDRSP-FL), addressing limitations in current research that restrict drone operations to customer sites only. Unlike restricted Launch and Recovery Operations (LARO) that rely primarily on truck deliveries and result in longer routes and underutilised drones, our model allows the truck to use non-customer locations (flexible sites) for drone operations, enhancing delivery efficiencies.
The proposed Mixed Integer Linear Programming (MILP) model makes three key decisions:
determining whether customer locations should be served by truck or drone; optimising truck and drone routing; and scheduling drone launch and recovery times at various truck stops. This flexibility in drone operation site selection facilitates faster deliveries and improved drone utilisation, particularly advantageous given drones' range limitations.However, the MILP model becomes computationally intractable for large instances. To address this, we developed a new branch-price and cut algorithm that improves scalability and minimises delivery times. Our findings indicate that using flexible sites overcomes the constraints of high-demand density locations and optimises the truck route, significantly enhancing the efficiency of the delivery system. This approach also results in higher drone utilisation compared to the traditional truck–drone setup. The model offers several practical insights for selecting potential flexible stops and the efficient management of a mixed fleet of trucks and drones for last-mile deliveries.

2 - Neural Combinatorial Optimization for Vehicle Routing Problems with Drones

Aigerim Bogyrbayeva, Cal Poly State, San Luis Obispo, CA, United States, Aigerim Bogyrbayeva

Many exact algorithms, heuristics, and metaheuristics have been proposed to solve the Vehicle Routing Problem with Drones (VRPD), which involves using a fleet of trucks and drones to fulfill customer orders in last-mile delivery. In this study, we formulate this problem using the Markov Decision Process (MDP) and propose a Reinforcement Learning (RL) based solution. Our RL model is based on an attention-encoder and a recurrent neural network-decoder architecture. This approach enhances coordination by determining which vehicles should visit specific customers and where vehicles can rendezvous, effectively leveraging drones and reducing the overall completion time. Our proposed RL model has demonstrated competitive performance compared

3 - Optimal Freight-on-Transit Operations Using Drones

Ouyang Yanfeng, University of Illinois Urbana-Champaign, Urbana, IL, United States, Jesus Osorio

This paper investigates the integration of Freight-on-Transit (FoT) with drone operations to optimize urban freight logistics. It develops a mathematical model for optimizing drone-assisted FoT operations, taking into account drone battery limitations, transit vehicle capacities, and the spatial-temporal alignment of transit and delivery schedules. A Lagrangian relaxation with column generation approach is proposed as a solution method to manage the computational complexity. Numerical experiments demonstrate the efficacy of this approach compared to commercial solvers and a simulated annealing metaheuristic. Sensitivity analyses highlight the impact of transit vehicle capacity and drone battery life on the overall system performance. A real-world case study using Chicago's Metra network illustrates practical applications and cost savings compared to truck-only and drone-only operations.

4 - Drone Scheduling Logistics

Srinivasa Prasanna, IIIT-Bangalore, Bengaluru, India

We describe drone scheduling algorithms suitable for agriculture, where crop conditions have to be identified, and appropriate fertilizers, pesticides, ... given. The problem is complex due to the scale and complex flying conditons, with uncertainty.

We desccribe both AIML heruistics and optimal methods. Results on simulated and real situations will be presented.

The worrk builds on earlier work on scheduling under uncertainty.

5 - Multi-Depot Drone Arc Routing Problem and Solution Methods

Islam Altin, Eskisehir Osmangazi University, Eskisehir, Turkey, Aydin Sipahioglu, ILGIN ACAR

Recently, drones have gained popularity in the fields of delivery, logistics, and surveillance operations. Researchers have focused on route planning to improve the effectiveness of drones in these fields. While several studies have been conducted on drone routing, a substantial percentage of them have addressed this problem as node-based routing. Only a small number of studies in the literature define this problem as edge-based routing. This study describes an edge-based drone routing problem by taking surveillance drones into account. In this case, there are multiple depots and more than one drone in each depot. Drones have limited energy capacity and must perform edge monitoring operations with their limited energy capacity. The objective of this problem is to obtain optimal drone routes that minimize the total distance traveled. However, it is not mandatory to use all the depots in the problem. Therefore, in addition to getting the optimal drone routes, it is necessary to determine which depots to use. We refer to this problem as the Multi-Depot Drone Arc Routing Problem. We developed a mixed-integer mathematical model to obtain the optimal solution to this challenging problem. Furthermore, we used a metaheuristic algorithm to handle especially large-sized problems in reasonable computation time. The algorithm's local search operators were specifically tailored to the problem. Real test problems of different sizes were generated in order to evaluate the performance of these solution methods. A comprehensive analysis of the computational results is presented.

6 - Optimizing Battery Recharging for Drones in Last-Mile Logistics Under Uncertainty: a Novel Decomposition-Based Approach

Ali Tolooie, Mercer University, Macon, GA, United States

This study addresses the operational challenges in drone logistics, focusing on the optimization of recharging stations, a vital but underexplored area. Our research aims to develop comprehensive strategies for efficient battery management, optimal charging times, and effective inventory control under variable conditions such as fluctuating demand and variable flight and charging times. Our methodology revolves around developing Markov decision process (MDP) models to optimize operations at drone recharging stations, taking into account a range of factors including battery charging strategies, inventory management, and dynamic operational conditions. Theoretically, we identify specific conditions that determine the optimal policy and characterize its structure. A key innovation in our approach is the introduction of a novel decomposition-based method to address the computational complexities of large-scale MDPs. This method segments the original problem into multiple independent processes, each corresponding to a different subsystem within the network, enhancing computational efficiency. Our numerical studies demonstrate the effectiveness of this approach, highlighting how various stochastic factors like flight patterns, demand variability, and inventory and backorder costs impact optimal decision-making. This research contributes to the field of drone logistics by providing actionable strategies for improved last-mile delivery, with broader implications for sustainable, cost-effective, and customer-focused logistics solutions.

SB14

Summit - 334

Financial Analytics and Technology

Invited Session

Finance

Chair: Agostino Capponi, Columbia University, New York, NY, United States

1 - A Distributionally Robust Instrumental Variable Estimation Framework

Zhaonan Qu, Stanford University, Stanford, CA, United States

We propose a distributionally robust optimization (DRO) formulation of the classical instrumental variable (IV) regression framework. When the ambiguity set is a Wasserstein ball centered at the empirical distribution of projected variables, the resulting estimator solves a square root version of the standard ridge regularized IV regression. We show that the estimator is consistent whenever the limit of the robustness/regularization parameter is bounded above by an estimable non-zero constant. This novel consistency result differs from existing ones on regularized regression because it allows a non-vanishing regularization parameter. We discuss some interesting features of the problem underlying this phenomenon, namely a projection-based ambiguity set coupled with the unique geometry of the square root ridge regression. We also discuss generalizations to other DRO problems with projection-based ambiguity sets.

2 - Optimal Exiting for Liquidity Provision in Constant Function Market Makers

Brian Zhu, Columbia University, New York, NY, United States, Agostino Capponi

Liquidity provision in exchanges that use constant function market makers (CFMMs) is often unprofitable or unfavorable relative to just holding assets, one major reason being impermanent loss. Using an optimal stopping approach, we show that under mild conditions on the pricing function, there exists an optimal upper threshold policy that yields positive returns in expectation to liquidity providers when they exit the liquidity pool once the price ratio hits the upper price threshold (before transaction and infrastructural fees). We also examine how the pricing function's curvature affects the degree of impermanent loss and the optimal upper threshold.

3 - Asset pricing with supply chain relationships

Jiacheng Zou, Stanford University, Menlo Park, CA, United States

We propose to leverage the comprehensive records of Bloomberg SPLC to study how supply chain (SC) information can help construct better asset pricing models. Asset pricing models explain correlation of stock performances and be used to construct portfolios with desirable properties, such as uncorrelatedness with the market. From the microeconomic perspective, SC data is fundamentally descriptive of firms' input-output relations in terms of production and sales and should contribute to explaining the correlations of their stock performances. However, as a gap between theory and empirics, there is a lack of studies that use SC data for asset pricing, due to 3 key challenges: (i) Large scale. Each year, there may be as high as tens of thousands of relationships in the SC of the SP500 constituent companies; (ii) Pervasive missingness. The SEC only requires disclosing relationships of over 10% revenue, and SC data also contains non-quantitative links; (iii) Lagging updates. Many SC disclosures appear only in annual reports and may dissolve without clear updates. Machine learning offers robust models to extract signals from noisy data, which has been used in the academic literature to construct asset pricing models with firm characteristics and showed promising results. Another gap is that so far the literature focuses on indirectly learning covariances of stocks from price co-movement, rather than directly from production and sales links, partially because the traditional time-series or cross-sectional dimensions of financial market data are easier to model. We aim to close both gaps by comprehensively studying relationships for asset pricing

SB15

Summit - 335

Trending Themes in Innovation and Development

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Gulru Ozkan-Seely, University of Washington, Bothell, WA, United States

1 - When Mission Drift Becomes Mission Adaptation for Social Ventures

Leonardo Santiago, Copenhagen Business School, Frederiksberg, Denmark, Sinan Erzurumlu, Juliana Hsuan, Moren Levesque

New ventures with the goal for solving a social problem often face an inherent dilemma: generating sustained profitability vs. generating social value. While a venture's commercial activities generate cash, the venture might be criticized for doing that as it can result in a mission drift whereby the venture's activities drift from directly achieving its social objectives over time. We explore the role of mission drift by developing a stylized welfare-maximization model that combines social value and profits. Our primary interest is in understanding the optimal timing of the mission drift that balances sustained profitability and social value generation.

2 - How do mandated mobile app updates influence user engagement and app performance?

Alejandro Hernandez De la Lanza, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Maria Ibanez

Mobile app developers are commonly believed to have complete discretion over the content they release through updates. However, mobile app marketplaces (i.e., Google Play, App Store) often require developers to release specific updates to ensure ecosystem compatibility. Failure to release them by specific deadlines can have consequences for developers: the app can lose visibility on the platform. In this study, we rely on a natural experiment research design and application of the difference-in-differences methodology to examine the impact of such mandatory updates, versus those at developers' discretion, on user behavior in a freemium app over time. We hope our findings can help developers implement timely updating strategies that minimize user base churn and allow for more efficient promotional campaigns.

3 - Artificial intelligence utilization and collaboration for new product development

Gulru Ozkan-Seely, University of Washington Bothell, Bothell, WA, United States

The utilization of artificial intelligence (AI) in the design and development of new products is rapidly advancing, particularly in technologically oriented fields. We investigate the impact of AI in design and development, focusing on the advantages and challenges of this practice. Specifically, we recognize that while AI can efficiently advance early design ideas or prototypes through task automation, processing large data sets, improving precision, and optimizing features and resource usage, it also poses hurdles. These include the complexity of AI integration, opaque information processing, overfitting to data, and limited creativity opportunities. We examine the trade-off between the benefits of using AI to improve designs and prototypes and the drawbacks associated with these challenges.

SB16

Summit - 336

Bora Keskin's RM&P Session (Prominent Students incl. Job Market Candidates)

Invited Session

Revenue Management and Pricing

Chair: Bora Keskin, Duke University, Durham, NC, United States

Co-Chair: Morgan Wood, UNC Chapel Hill, Whitakers, NC, United States

1 - A Discretization Framework for Robust Contextual Stochastic Optimization

Cristian Rares, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgia Perakis

In revenue management, businesses often face decision-making challenges based off of optimization problems with uncertain, context-dependent random variables (such as demand). Due to the inherent uncertainty in these problems, one is often interested not only in minimizing expected cost, but also to be robust and protect against worst case scenarios. We propose a novel method that combines the learning stage with knowledge of the downstream optimization task. The method prescribes decisions which aim to maximize the likelihood that the cost is below a (user-controlled) threshold. We provide theoretical guarantees bounding the regret associated with decisions proposed by our method. Furthermore, experimental results demonstrate the competitiveness and robustness of our approach. In particular, our method yields up to 20 times lower worst-case revenue loss compared to other methods in a real-world scenarios, such as electricity generation or inventory allocation, while maintaining comparable average revenue performance.

2 - Platform Competition in Two-Sided Networks

Emin Ozyoruk, University of Chicago, Chicago, IL, United States, John Birge

We study the impact of competition on ride-hailing platforms. The riders are sensitive to price and delay, and the drivers value high wages and utilization. The efficiency of the platforms improves with the density, and the drivers are randomly located in relation to the riders. The platforms simultaneously choose prices, service standards, and wages. The game admits a tractable form where each platform chooses a full price, enabling us to characterize the equilibrium and establish its existence and uniqueness. The results suggest that the ride-hailing industry benefits from economies of scale as the unit cost of attracting drivers decreases with increasing demand. Multiple platforms can enter and co-exist due to the multihoming drivers and variability in the system. However, a limited number of platforms can achieve economies of density and be profitable.

3 - Data-Driven Dynamic Assortment in Online Platforms: Learning about Two-Sides

Rahul Roy, Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Nur Sunar, Jayashankar Swaminathan

Unlike traditional marketplaces like Amazon, a two-sided online platform allows "active" participants, i.e., they must initiate proposals to interact with agents on the other side to facilitate transactions. With the rise of such platforms (e.g., HomeAdvisor), understanding participants' preferences plays a crucial role in enhancing platforms' financial performance. However, these preferences are unknown to the platform, and must be learned over time. Furthermore, platforms often lack adaptive decision support systems that facilitate transactions between participants under uncertainty. This paper aims to bridge this crucial gap by formulating a dynamic assortment selection problem faced by a two-sided platform, where heterogeneous, active customers sequentially arrive to acquire products, and the platform chooses a subset of sellers from a finite set, tailored to the customer type. Each customer, being active, sends a proposal to at most one seller to initiate a transaction. After recurring intervals, all sellers review their proposals and matches with at most one customer. Platform aims to learn participants' preferences while maximizing the expected reward over time. Unlike extant literature that has solely analyzed one-way learning scenarios where the focus is to learn only customer preferences, our study proposes a novel formulation: "two-way learning." In our study, we develop an asymptotically optimal online algorithm that dynamically learns participants' preferences while optimizing platform's objective. Our algorithm achieves a worst-case regret bound that significantly improves upon the regret bounds in the existing literature. Our study contributes to the advancement of assortment optimization online algorithms, especially within the framework of two-way learning.

4 - Data-driven Population Tracking in Large Service Systems

Morgan Wood, UNC Chapel Hill, Chapel Hill, NC, United States

We develop asymptotically optimal policies to track queue lengths under different loss structures in a setting with inaccurate arrival and departure sensor data. By deriving general lower bounds on any policy's expected cumulative loss, we show that tracking the population in real-time is challenging because of the accumulation of tracking errors. We propose a detection policy with provably good performance guarantees that corrects the estimated population count when the system is believed to be nearly empty. Further, we explore the benefits of augmenting the sensor information with periodic queue inspections at a cost. Our model is motivated by queue tracking implemented at a large airport. (https://ssrn.com/abstract=4748063)

SB17

Summit - 337

RMP in the Supply Chain Context

Invited Session

Revenue Management and Pricing

Chair: Levi DeValve, Chicago Booth, Chicago, IL, United States

1 - Multi-echelon Batch Inventory Networks

Maximiliano Stock, University of Chicago, Chicago, IL, United States, Levi DeValve

We consider the Multi-echelon Batch Inventory (MEBI) problem, which integrates the challenges of optimizing pallet ordering, breaking down these pallets into inventory units, and allocating these units across a network in anticipation of uncertain demand. We aim to minimize the combined costs of ordering, breaking down, placing inventory, and the expected costs associated with fulfillment and shortages. The research unveils the intrinsic complexity of the problem, demonstrating its inapproximability in terms of lower bounds under non-metric cost assumptions. Under the metric cost assumption, the usual LP relaxation for the two-stage stochastic optimization problem exhibits an unbounded integrality gap. To address this, we propose an improved linear programming relaxation of the two-stage stochastic optimization problem by incorporating additional cuts that help bound the fractional results that may arise from the relaxation of the integer constraints. We develop a constant factor approximation guarantee by implementing a randomized filtering rounding algorithm inspired by approximation algorithm techniques from the facility location literature. The findings enhance the theoretical understanding of the MEBI problem and offer practical insights for large-scale operations in optimizing complex inventory systems under uncertainty.

2 - Landscape of Policy Gradient Objectives for Finite Horizon MDPs: Applications in Operations Models

Minda Zhao, Georgia Institute of Technology, Atlanta, GA, United States, Xin Chen, Yifan Hu

We explore policy gradient methods for finite horizon Markov Decision Processes (MDPs) with continuous state and action spaces. Policy gradient methods for Markov Decision Processes (MDPs) do not converge to global optimal solutions in general due to the non-convexity of the objective functions. We identify several easily verifiable conditions to guarantee the global Kurdyka-Łojasiewicz (KŁ) condition for the objectives of policy gradient optimization problems for a class of MDPs. This allows us to establish that the policy gradient optimization problems can be solved by first-order methods with a sample complexity in the order of $1/\varepsilon$ and polynomial at the planning horizon length to attain ε-global optimal solutions. Our results find applications in a host of operations models including the stochastic cash balance problem and multi-period inventory system with Markov-modulated demand, giving the first sample complexity results in the literature.

3 - Private Optimal Inventory Policy Learning for Feature-based Newsvendor with Unknown Demand

Tuoyi Zhao, University of Miami, Miami, FL, United States, Wenxin Zhou, Lan Wang

The data-driven newsvendor problem with features has recently emerged as a significant area of research, driven by the proliferation of data across various sectors such as retail, supply chains, e-commerce, and healthcare. Given the sensitive nature of customer or organizational data often used in feature-based analysis, it is crucial to ensure individual privacy to uphold trust and confidence. Despite its importance, privacy preservation in the context of inventory planning remains unexplored.
A key challenge is the nonsmoothness of the newsvendor loss function, which sets it apart from existing work on privacy-preserving algorithms in other settings. This paper introduces a novel approach to estimate a privacy-preserving optimal inventory policy within the f-differential privacy framework, an extension of the classical (ϵ, δ)-differential privacy with several appealing properties.
We develop a clipped noisy gradient descent algorithm based on convolution smoothing for optimal inventory estimation to simultaneously address three main challenges: (1) unknown demand distribution and nonsmooth loss function; (2) provable privacy guarantees for individual-level data; and (3) desirable statistical precision.
We derive finite-sample high-probability bounds for optimal policy parameter estimation and regret analysis.
By leveraging the structure of the newsvendor problem, we attain a faster excess population risk bound compared to that obtained from an indiscriminate application of existing results for general nonsmooth convex loss. Our bound aligns with that for strongly convex and smooth loss function.

4 - Beyond Basic Reusability: Joint inventory and online assortment optimization with a network of evolving resources

Yukai Huang, Olin Business School, St. Louis, MO, United States, Jacob Feldman, Heng Zhang

In this study, we consider the joint inventory and assortment optimization when the decision-makers decide the initial capacity of each resource before a finite selling season starts. Then, across the selling horizon, the decision-makers offer personalized assortment in real time based on customer arrivals and remaining inventory to maximize the total expected revenue. Unlike prior work on joint inventory selection and online resource allocation that only considers the perishable resources, our framework can be extended to the reusable resources. Beyond the classic definition of reusable resource, we introduce the new concept of network reusability by allowing for the transformation of resources upon return. This advanced model captures more realistic scenarios, such as products returning to inventory in a different condition or rentals returning to different locations, thereby broadening the applicability of our framework to diverse domains like product returns, fashion rentals, and car-sharing services. The modeling richness introduces new technical hurdles when it comes to solving our joint inventory and online assortment problem. Under the MNL choice model, we propose an inventory refinement procedure that achieves a constant-factor approximate performance. Our computational experiments show that our approximation framework performs well under different application scenarios.

SB18

Summit - 338

Revenue Management and Online Platform Operations

Invited Session

Revenue Management and Pricing

Chair: Can Kucukgul, Rutgers University - Camden, Philadelphia, PA, United States

1 - Unveiling Competitive Dynamics in Pricing: Insights from Laboratory Experiments

Bahriye Cesaret, Ozyegin University, Istanbul, Turkey, Armagan Bayram

We conduct a series of controlled laboratory experiments to investigate the dynamic pricing behavior of two capacity-constrained firms under competition. Our experiments initially control for strategic interactions between the sellers and then allow for them. Moreover, we manipulate demand uncertainty and the expected market size to broaden our comprehension. Our findings underscore the significant influence of competitor's behavior on dynamic pricing decisions. We find that the theory predicts a lower level of competition among the sellers than what we observe in the laboratory. The modeling literature indicates that the seller with the lower capacity has a competitive advantage, but our results reveal the opposite. Further, there is potential for high-capacity sellers to benefit from competition. Sellers tend to underprice (resp., overprice) their units at the beginning (resp., end) of a selling season. Also, the duration of competition exceeds theoretical predictions, and customers reap advantages from the biases inherent in strategies of competing sellers. Leveraging our experimental findings, we propose two straightforward heuristics that utilize the competitor's past behavior to improve predictions of sellers' pricing behavior over the normative benchmarks.

2 - Optimal Assortment and Free Shipping Strategies: a Comparative Analysis

Hafila Morais Max Morais, Texas State University, McCoy College of Business, San Marcos, TX, United States, Sahika Sahan Konur

This paper explores optimal assortment decisions of a retailer across various free shipping strategies. Utilizing real-world data, we conduct a computational analysis to compare the expected revenues under two distinct scenarios: one where the retailer incurs all the shipping costs (free shipping), and another where the customer incurs the entire shipping cost (no free shipping). Our study aims to illuminate the impact of different shipping strategies on retailer's revenue and decision-making processes.

3 - The Value of Information Design in Supply Chain Management

Huseyin Gurkan, ESMT GmbH, Berlin, Germany, Ozan Candogan

This paper studies an information design problem of a retailer in a two-tier supply chain that procures a single type of product from a supplier. The supplier needs to decide on a production quantity by balancing the shortage cost and the excess inventory holding cost with respect to the retailer's demand. The retailer's demand is random but the retailer receives an informative signal about the demand before the supplier sets the production quantity, and places orders after learning the demand realization. The retailer wants to reduce the shortage cost, and to this end she can disclose information about her signal to persuade the supplier to increase production levels. For this setup, we characterize the optimal information disclosure policy of the retailer, and shed light on settings where the retailer strictly benefits from carefully designed information disclosure policies relative to a full- or a no-disclosure policy.

4 - Subscription Pricing Strategies for Delivery Services

Shankar Sundaresan, Rutgers University, Camden, NJ, United States, Anantaram Balakrishnan

We study the subscription pricing problem that offers consumers two delivery choices: a pay-per-delivery option and a subscription option with free delivery. We develop and analyze a model that captures consumer heterogeneity in terms of their utility and preference across different firms, lift in consumer demand, possibility of attracting new customers, and order batching by consumers. We characterize the retailer's optimal subscription pricing strategies, and develop interesting insights.

5 - Dynamic pricing with random yields

Srinivasa Puranam, Rutgers University, Camden, NJ, United States, Michael Katehakis

This talk explores the impact of uncertainty surrounding realized capacity on a firm's pricing decisions, particularly in markets where contracts are agreed upon before the actual product delivery. In such markets, if the actual capacity realized at the end of the selling period falls short of the contracted capacity, the seller must adjust existing contracts and may not be able to meet the demand of one or multiple buyers in full. Many firms encounter exogenous factors that influence the final yield of their products, making it challenging to accurately predict yield. Consequently, firms often rely on estimates when selling capacity in advance of their actual yield, which can result in either shortfalls or excesses at the end of the selling season. We present a Markov Decision Process model for this problem.

SB19

Summit - 339

Mechanism Design Under Uncertainty

Invited Session

Revenue Management and Pricing

Chair: Shixin Wang, The Chinese University of Hong Kong, Hong Kong, N/A

1 - From Doubt to Devotion: Trials and Learning-Based Pricing

Nicholas Wu, Yale University, New Haven, CT, United States, Tan Gan

An informed seller designs a dynamic mechanism to sell an experience good. The seller has partial information about the product match, which affects the buyer's private consumption experience. We characterize equilibrium mechanisms of this dynamic informed principal problem. The belief gap between the informed seller and the uninformed buyer, coupled with the buyer's learning, gives rise to mechanisms that provide the skeptical buyer with limited access to the product and an option to upgrade if the buyer is swayed by a good experience. Depending on the seller's screening technology, this takes the form of free/discounted trials or tiered pricing, which are prevalent in digital markets. In contrast to static environments, having consumer data can reduce sellers' revenue in equilibrium, as they fine-tune the dynamic design with their data forecasting the buyer's learning process.

2 - Pricing with Binary Feedback: the Value of Information Bit by Bit

Omar Mouchtaki, New York University, New York, NY, United States, Omar Besbes, Will Ma

We study a data-driven pricing problem in which a seller offers a price for a single item based on demand observed at a small number of historical prices. Our goal is to derive precise evaluation procedures of the value of the historical information gathered by the seller, along with prescriptions for more efficient price experimentation. Our main methodological result is a reduction which allows us to compute exactly the worst-case performance of a wide range of pricing policies for any sequence of historical prices. This allows us to establish that very few demand observations are sufficient to ensure a near-optimal revenue.

3 - The Power of Simple Menus in Robust Selling Mechanisms

Shixin Wang, The Chinese University of Hong Kong, Hong Kong, Hong Kong

We study the robust screening problem with a finite menu size. Our framework characterizes the optimal selling mechanisms and the corresponding competitive ratio across various menu sizes and ambiguity sets of the buyer's valuation distributions, such as support, mean, and quantile ambiguity sets. We show that a selling mechanism with a small menu already yields a competitive ratio comparable to the optimal robust mechanism with infinite options in the menu. Remarkably, a menu size of merely two can significantly enhance the competitive ratio compared to deterministic pricing, which establishes a favorable trade-off between theoretical performance and implementation simplicity.

SB20

Summit - 340

Human-Centered AI and Decision Analysis for Social Good

Invited Session

Decision Analysis Society

Chair: Yunyi Li, The University of Texas at Austin, Austin, TX, United States

1 - A Methodological Framework for Human-Algorithm Performance Comparisons Under Uncertainty

Luke Guerdan, Carnegie Mellon University, Pittsburgh, PA, United States, Amanda Coston, Steven Wu, Kenneth Holstein

Predictive models are often introduced to decision-making tasks under the rationale that they improve performance over an existing (e.g., human) decision-making policy. However, it is challenging to compare predictive performance against an existing decision-making policy that is generally under-specified and dependent on unobservable factors. These sources of uncertainty are often addressed in practice by making strong assumptions about the data-generating mechanism. In this work, we propose a method to compare the predictive performance of decision policies under a variety of modern identification approaches from the causal inference and off-policy evaluation literatures (e.g., instrumental variable, marginal sensitivity model, proximal variable). Key to our method is the insight that there are regions of uncertainty that we can safely ignore in the policy comparison. We develop a practical approach for finite-sample estimation of regret intervals under no assumptions on the parametric form of the status quo policy. We verify our framework theoretically and via synthetic data experiments. We conclude with a real-world application using our framework to support a pre-deployment evaluation of a proposed modification to a healthcare enrollment policy.

2 - Mitigating Label Bias via Decoupled Confident Learning

Yunyi Li, The University of Texas at Austin, Austin, TX, United States

Growing concerns regarding algorithmic fairness have led to a surge in methodologies to mitigate algorithmic bias. However, such methodologies largely assume that observed labels in training data are correct. This is problematic because bias in labels is pervasive across important domains, including healthcare, hiring, and content moderation. In particular, humangenerated labels are prone to encoding societal biases. While the presence of labeling bias has been discussed conceptually, there is a lack of methodologies to address this problem. We propose a pruning method—Decoupled Confident Learning (DeCoLe)—specifically designed to mitigate label bias. After illustrating its performance on a synthetic dataset, we apply DeCoLe in the context of hate speech detection, where label bias has been recognized as an important challenge, and show that it successfully identifies biased labels and outperforms competing approaches.

3 - Does a Human-Algorithm Feedback Loop Lead To Error Propagation? Evidence from Zillow's Zestimate

Runshan Fu, New York University, New York, NY, United States, Ginger Jin, Meng Liu

We study how home sellers and buyers interact with Zillow's Zestimate algorithm throughout the sales cycle of residential properties, with an emphasis on the implications of such interactions. In particular, leveraging Zestimate's algorithm updates as exogenous shocks, we find evidence for a human-algorithm feedback loop: listing and selling outcomes respond significantly to Zestimate, and Zestimate is quickly updated for the focal and comparable homes after a property is listed or sold. This raises a concern that housing market disturbances may propagate and persist because of the feedback loop. However, simulations suggest that disturbances are short-lived and diminish eventually, mainly because all marginal effects across stages of the selling process—though sizable and significant—are less than one. To further validate this in- sight in the real data, we leverage the COVID-19 pandemic as a natural experiment. We find consistent evidence that the initial disturbances created by the March-2020 declaration of na- tional emergency faded away in a few months. Overall, our results identify the human-algorithm feedback loop in an important real-world setting, but dismiss the concern that such a feedback loop generates persistent error propagation.

4 - A Machine Learning-Based Approach for Evaluating Experts' Decision Quality in High-Risk Instances

Wanxue Dong, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Weiquan Wang

Experts are fallible, and the repercussions of their decisions can be substantial. For instance, misdiagnoses by physicians in the US annually result in billions of dollars in losses and impact millions of lives. Hence, assessing experts' decisions is vital for guiding managerial tasks and consumer services. However, evaluating decision quality is challenging due to the scarcity and high cost of obtaining ground truth (GT) decisions—those that are correct. While previous studies have attempted to assess experts' decision quality with scarce GT, they often hinge on the assumption that GT and experts' decisions stem from the same distribution. However, such assumptions overlook crucial real-world scenarios. Advanced treatments or medical examinations to uncover GT, for instance, may only be accessible to patients deemed high-risk. Therefore, prior solutions assessing decisions in general may not adequately address the needs of high-risk groups, which are also more susceptible to decision errors. In this study, we concentrate on scenarios where GT data is limited and only accessible for instances with a higher risk of errors. Thus, we propose a Machine Learning (ML)-based approach to assess experts' decision quality in high-risk instances by harnessing a large amount of experts' noisy decisions and scarce GT data from high-risk cases. Our empirical evaluations demonstrate that our method outperforms an alternative that rely solely on independent GT sets with high-risk instances. To our knowledge, this paper is the pioneer in considering high-risk decision evaluation with limited GT available from high-risk groups and achieves state-of-the-art performance.

5 - Elucidating Justice Concerns of Algorithms that Triage Misinformation for Fact-Checkers

Terrence Neumann, University of Texas at Austin, Austin, TX, United States, Maria De-Arteaga, Maytal Saar-Tsechansky, Matt Lease, Sina Fazelpour

Faced with the increasing scale of misinformation, fact-checkers are frequently turning to algorithms to efficiently triage claims in need of verification. However, there is uncertainty regarding the appropriate 'ground truth' for training and evaluating these algorithms given the varied and subjective factors that influence how claims are prioritized for checking. For instance, numerous fact-checking organizations prioritize checking claims that are most likely to go viral and cause harm to the 'general public,' while others have noted the need to prioritize claims that harm vulnerable demographic groups. To better understand the implications of these and other algorithmic design choices on society, we apply the theoretical lens of informational justice to elucidate the often-competing interests of representation, participation, credibility, and the distribution of benefits and burdens among the stakeholders affected by information proliferation, consumption, and fact-checking. Utilizing a new crowdsourced dataset, which we make available for public use, we apply the informational justice framework to a training data audit. We show how algorithmic design choices related to the construction of a 'ground truth' label for claim prioritization can have consequential justice implications for stakeholders affected by misinformation. We propose that an informational justice approach promotes transparency and facilitates more ethical decision-making amidst the unique complexities of claim triage and prioritization of fact-checking resources. Our findings aim to assist researchers, policymakers, and industry professionals in framing the ethical landscape of AI-assisted fact-checking systems.

SB21

Summit - 341

Reinventing Business Models Amidst New Market Dynamics

Invited Session

Decision Analysis Society

Chair: Yuan Cheng, Cornell University, Jersey City, NJ, United States

1 - Selling Only? Leasing Only? Maybe Selling and Leasing Combined?

Shuya Yin, University of California at Irvine, Irvine, CA, United States, Mobin Ghasempour Nejati

The selling and leasing business models in the automobile industry have been extensively used in practice and studied in literature. With the rising prevalence of electric vehicles, with battery components being a significant part of these products and their depreciation rate being higher than other parts, it seems natural to explore a combined business model where the battery can be leased or rented while the vehicle (excluding the battery) can be sold as usual. This paper examines the trade-offs involved in such considerations.

2 - Fair Contract Design with Agent Rotation

Feng Tian, HKU, Hong Kong, China, People's Republic of, Mingliu Chen, Wei Zhang

We study a dynamic optimal contracting problem when agents have fairness concerns about their earnings. A principal hires agents to run businesses in different markets with different costs. Agents can exert costly private efforts to potentially increase the chance of business success in their markets, and the principal only observes the outcome of the businesses. All agents and the principal are risk-neutral, but the agents have fairness concerns about their earnings due to market differences. The principal can use compensations and agent rotation to alleviate agents' concerns about fairness. At the same time, the principal also wants to induce efforts from the agents to increase the chance of business success and maximize her utility. Note that when subjected to fairness concerns, the principal indirectly provides the agents with some compensation or utility guarantees. As a result, it may be more difficult or costly to motivate agents to exert efforts as the monetary incentives are less effective. On the other hand, a rotation schedule may also serve as an effective incentive or threat to agents, depending on the market conditions, but it may lead to costly logistics. As a result, the principal needs to properly design dynamic contracts for agents involving compensations and agent rotation based on the business outcomes over time to carefully balance their incentives to work hard and fairness concerns.

3 - Using Semantic Network Analysis to Measure the Codification of Occupational Ethics

Jessica Santana, UC Santa Barbara, Goleta, CA, United States, Seonghoon Kim

The institutionalization of occupations tends to assume homogenization of occupational values. This study addresses the question of how members of an occupation with dissenting preferences reach consensus on a code of ethics. We build on prior theorization of occupational institutionalization and institutional discourse to theorize ethical codification as a dynamic discursive process of internal dissent and consensus culminating in a professional code of ethics. We use email data from the IEEE-ACM Software Engineering Ethics and Professional Practice Committee tasked with producing the 1997 Software Engineering Code of Ethics to show how ethical codification follows a process of initial competition followed by semantic convergence. This study demonstrates how natural language processing and semantic network analysis can contribute to discourse analyses of institutional processes.

SB22

Summit - 342

Wisdom of the Crowd/Forecast Aggregation

Invited Session

Decision Analysis Society

Chair: Zhi Chen, National University of Singapore, Singapore, Singapore

Co-Chair: Long Zhao, NUS DAO, Singapore, Singapore

1 - Bayesian Ensembles of Unlabeled Forecasts: a Rationale for Downweighting Extreme Forecasts

Xiaojia Guo, Robert H. Smith School of Business, University of Maryland, College Park, MD, United States, Casey Lichtendahl

We introduce an ensemble method based on a Bayesian framework that incorporates forecasts and biases from various experts or models. The final uncertainty in the model removes exact knowledge about which expert is best, second best, etc., which in practice is often the case. The final model, then, learns only from the order statistics of the forecasts, as if the forecasts were unlabeled. In other words, the model learns anonymously, without knowing the identity of the experts or how they performed in the past. According to the model, the more extreme a forecast is, the more likely it comes from an overfit or underfit expert or model, and therefore it receives less weight in the ensemble. In two empirical studies of time series forecasts, we demonstrate that our Bayesian ensemble can outperform leading methods.

2 - Regularized Aggregation of Point Predictions from Experts with Different Amounts of past Performance Data

Junnan Wang, INSEAD, Singapore, Singapore, Ville Satopaa

Aggregating point forecasts from multiple experts is a critical challenge, especially in the common context where those experts have contributed a varied number of predictions in the past. Consider the scenario where one expert has provided just two forecasts but with perfect accuracy, and another has offered a hundred predictions with lower accuracy. This situation presents a challenge in balancing accuracy against the risk associated with less data. To address this, our work introduces a novel aggregation method that precisely balances these considerations. Specifically, our aggregator uses Bayesian regularization, estimates each expert's expected bias and noise, factoring in their unique prediction histories, and appropriately shrinks each expert's weight toward equal weight, depending on the expert's skill and how reliably that skill can be estimated. In our empirical study, we apply our aggregator to the data from the European Central Bank's Survey of Professional Forecasters and observe that it outperforms traditional aggregation methods, including the simple average and the performance-based weighted average methods.

3 - From Point Predictions to Probabilistic Aggregate: A Novel Variational Bayes Approach to Harnessing the Wisdom of the Crowd.

Yanwei Jia, The Chinese University of Hong Kong, Shatin, Hong Kong, Jussi Keppo, Ville Satopaa

Decision makers often rely on the combination of many individual judgments, which is known as the wisdom of crowds. In the single-question setting, where there is no past data or other prior information, recent literature finds that using peer predictions, i.e., predictions about others' judgments, can greatly enhance the accuracy of aggregated judgment. However, the methods proposed so far either request individuals provide probabilistic predictions or are restricted to output only an aggregated single-point prediction. The former is difficult to elicit and the latter is not sufficient to inform the decision-making. By using variational Bayes inference techniques, we propose an aggregation method that takes the individual judgments and their peer predictions, both single-point values, as inputs to construct a probability distribution that can be expressed in an analytical form. Under Normal-Gamma information structure, we prove the efficiency of the proposed method. Using simulated data and experimental data collected in previous studies, we demonstrate the proposed method outperforms conventional practices and achieves comparable accuracy to state-of-the-art methods in single-point prediction.

4 - "Information Elicitation from Teams of Privacy-Conscious Experts

Marat Salikhov, New Economic School, Moscow, Russian Federation, Ruslan Momot

Companies often inform their decisions by eliciting votes from teams of internal experts – for example, their employees. However, such experts may vote against their true beliefs if pressured by powerful stakeholders, whose interests may be misaligned with the company's best interest. The experts' desire to protect their true beliefs from such stakeholders presents a privacy concern and makes eliciting their information a challenge. In particular, our results reveal that privacy concerns may undermine the conventional "wisdom of crowds" logic –

in which having a larger pool of experts strictly improves the quality of the firm's decisions. We explore how privacy concerns can be addressed by designing appropriate vote elicitation processes.

5 - Combining Forecasts from Multiple Experts for Multiple Variables

Long Zhao, NUS DAO, Singapore, Singapore, Zhi Chen

We address the challenge of combining point forecasts from multiple experts across multiple variables to draw inferences about an unknown variable. A significant amount of prior research has focused on aggregating forecasts from multiple experts about a single variable into a consensus forecast. However, in practice, each expert frequently provides a forecast for each variable across multiple variables. In this paper, we study how a decision-maker can incorporate forecasts of other variables in addition to those provided for the focal variable.

SB23

Summit - 343

The MCDM Junior Researcher Best Paper Award Session

Award Session

Multi Criteria Decision Making

Chair: Salvatore Greco, University of Catania, Catania, Italy

Co-Chair: Adiel De Almeida, Universidade Federal de Pernambuco, Recife, Brazil

1 - On solving parametric multiobjective quadratic programs with parameters in general locations

Andrew Pangia, UNC Charlotte, Concord, NC, United States, Pubudu Jayasekara

While theoretical studies on parametric multiobjective programs (mpMOPs) have been steadily progressing, the algorithmic development has been comparatively limited despite the fact that parametric optimization can provide a complete parametric description of the efficient set. This paper puts forward the premise that parametrization of the efficient set of nonparametric MOPs can be combined with solving parametric MOPs because the algorithms performing the former can also be used to achieve the latter.This strategy is realized through (i) development of a generalized scalarization, (ii) a computational study of selected parametric optimization algorithms, and (iii) applications in a real-life context. Several variants of a generalized weighted-sum scalarization allow one to scalarize mpMOPs to match the capabilities of algorithms. Parametric multiobjective quadratic programs are scalarized into parametric quadratic programs (mpQPs) with linear and/or quadratic constraints. In the computational study, three algorithms capable of solving mpQPs are examined on synthetic instances and two of the algorithms are applied to decision-making problems in statistics and portfolio optimization. The real-life context reveals the interplay between the scalarizations and provides additional insight into the obtained parametric solution sets.

2 - An active preference learning approach to aid the selection of validators in blockchain environments

Grzegorz Miebs, Poznan University of Technology, Poznan, Poland, Jonas Gehrlein, Matteo Brunelli, Milosz Kadzinski

We consider a real-world problem faced in some blockchain ecosystems that select their active validators—the actors that maintain the blockchain—from a larger set of candidates through an election- based mechanism. Specifically, we focus on Polkadot, a protocol that aggregates preference lists from an- other set of actors, nominators, that contain a limited number of trusted validators and thereby influence the election's outcome. This process is financially incentivized but often overwhelms human decision makers due to the problem's complexity and the multitude of available alternatives. This paper presents a decision support system (DSS) to help the nominators choose the validators in an environment with frequently changing data. The system structures the relevant multiple attribute problem and incorporates a dedicated active learning algorithm. Its goal is to find a sufficiently small set of pairwise elicitation questions to infer nominators' preferences. We test the proposed solution in an experiment with 115 real nominators from the Polkadot ecosystem. The empirical results confirm that our approach outperforms the unaided process in terms of required interaction time, imposed cognitive effort, and offered efficacy. The developed DSS can be easily extended to other blockchain ecosystems.

3 - Cluster ensemble selection and consensus clustering: A multi-objective optimization approach

Dilay Aktas, KU Leuven Institute for Mobility, Leuven, Belgium, Banu Lokman, Tulin Inkaya, Gilles Dejaegere

Cluster ensembles have emerged as a powerful tool to obtain clusters of data points by combining a library of clustering solutions into a consensus solution. In this paper, we address the cluster ensemble selection problem and design a multi-objective optimization-based solution framework to produce consensus solutions. Given a library of clustering solutions, we first design a preprocessing procedure that measures the agreement of each clustering solution with the other solutions and eliminates the ones that may mislead the process. We then develop a multi-objective optimization algorithm that selects representative clustering solutions from the preprocessed library with respect to size, coverage, and diversity criteria and combines them into a single consensus solution, for which the true number of clusters is assumed to be unknown. We conduct experiments on different benchmark data sets. The results show that our approach yields more accurate consensus solutions compared to full-ensemble and the existing approaches for most data sets. We also present an application on the customer segmentation problem, where our approach is used to segment customers and to find a consensus solution for each segment, simultaneously.

SB24

Summit - 344

Corporate Governance and Firm Performance

Contributed Session

Contributed

Chair: Fei Li, The Hong Kong Polytechnic University

1 - An Empirical Analysis of Peer Firms Investigates the Relationship Between Corporate Governance Practices and Firm Performance

Tabreez Humayun, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Shuangyan LI, Shahzad Humayun, Gul Rukh

This study demonstrates that peer firms in Pakistan adopt the corporate governance practices that are currently being implemented. Additionally, it investigates the correlation between corporate governance procedures, capital structure, and firm performance in listed companies in Pakistan. Financial resources are essential for companies to achieve their aims. Consequently, we meticulously arrange the elements that can influence financing decisions for all peer firms. In this study, we utilize a sample of 100 peer organizations that are listed on the Pakistan Stock Exchange (PSE) as part of the non-financial industrial sector. The data covers the period from 2015 to 2020. The relationship between corporate governance, capital structure, and business performance is analyzed using Structural Equation Modeling (SEM) and regression approaches. The research study further substantiates the corporate governance ideas. The reliability of our conclusions is ensured by the use of theoretical models, which promote excellent corporate governance practices and ultimately improve the firm's performance and capital structure. The study examines the significance of corporate governance procedures in influencing peer firms to modify their debt strategies and improve overall firm performance. Furthermore, in order to acquire additional resources and cultivate trust from shareholders, organizations must enhance their corporate governance methods.

Keywords: Corporate Governance, CEO Duality, Debt-to-Equity Ratio, ROA, and

Earnings per Share.

2 - Will Diversity, Equity, and Inclusion Commitment Improve Manufacturing Firms' Market Performance? A Signaling Theory Perspective on DEI Announcements

Fei Li, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Chris K. Y. Lo, Yi Zhou, Christopher Tang

Despite the escalating emphasis on diversity, equity, and inclusion (DEI), there remains a noticeable gap in empirical research concerning its implications, particularly within the manufacturing sector. This research gap has prompted us to scrutinize DEI commitment's effects on publicly traded manufacturing corporations through the lens of *signaling theory*. To operationalize our study, we employ the event study methodology and apply the structural topic modeling method to analyze 233 DEI commitment announcements issued by 161 firms over a 10-year period between 2013 and 2022. Our findings suggest that DEI commitment announcements can yield positive abnormal stock returns during the announcement period (from day -1 to 0 and from day -1 to 5). Furthermore, the impact will be heightened when the announcements place a stronger emphasis on topics related to DEI as opposed to topics not related to DEI. This is referred to as *signal strength*. The effect is also more pronounced when the focus is on specific subjects within DEI, a concept we term *signal specificity*. We also discuss managerial implications of our findings.

SB25

Summit - 345

PSOR Best Paper Award Finalists

Award Session

Public Sector OR

Chair: Auyon Siddiq, University of California-Los Angeles, Los Angeles, CA, United States

Co-Chair: Jingwei Zhang, Cornell University, Ithaca, United States

1 - Improving the Equity and Efficiency of Cystic Fibrosis Screening

Douglas Bish, University of Alabama, Northport, AL, United States, Hussein El Hajj, Ebru Bish

Coming soon

2 - Optimizing Health Supply Chains with Decision-Aware Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Hamsa Bastani, Osbert Bastani

We address the problem of allocating limited medical resources in a developing country by combining ML (to predict demand) with optimization (to optimize allocations). A key challenge is aligning the ML model's loss function with the decision loss in optimization. We propose a scalable decision-aware learning framework and successfully deployed in collaboration with the Sierra Leone government across 1,123 healthcare facilities nationwide, leading to a 15-29% increase in medicine consumption and improved real-world patient access to care.

3 - Redesigning VolunteerMatch's Ranking Algorithm: Toward More Equitable Access to Volunteers

Akshaya Suresh, RAND Corporation, Washington, DC, United States, Vahideh Manshadi, Scott Rodilitz, Daniela Saban

Coming soon

4 - Search and Matching for Adoption from Foster Care

Vincent Slaugh, Cornell University, Ithaca, NY, United States, Ludwig Dierks, Nils Olberg, Sven Seuken, Utku Unver

Using a game theory model and outcomes from an agency in Florida, we analyze two paradigms for how to find families for children in need of adoption. In the existing family-driven search process, agencies announce available children to which families can express interest. In the new caseworker-driven search process, caseworkers evaluate families using a ranked list. We show the caseworker-driven approach can reduce search costs and lead to a higher likelihood of adoption for children.

5 - The Dedicated Docket in US Immigration Courts: An Analysis of Fairness and Efficiency Properties

Daniel Freund, MIT, Cambridge, MA, United States, Wentao Wang

coming soon

SB26

Summit - 346

Recent Developments in Innovation Contests

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Gizem Korpeoglu, TU/e, Eindhoven, 5632TE, Netherlands

Co-Chair: Konstantinos Stouras, UCD Michael Smurfit GSB, Dublin, Ireland

1 - Technology Contests

Jurgen Mihm, Insead, Fontainebleau, France, Jussi Keppo, Jochen Schlapp

When innovating companies often face the choice of which technology to implement in their product. In particular, firms can typically choose between an established or a more experimental technology. Which one should they choose? And when? And more interestingly, does competition change the optimal choice? And is competition thus good for society? In this paper, we build a formal model of a dynamic technology contest and we study the technology choices of competing firms over time. We find that a firm's technology choice mainly depends on the firm's relative position in the contest as well as the uncertainty embedded in experimenting with a novel technology. We also investigate whether such technology contests are desirable from a societal perspective: Do technology contests promote or hinder innovation? The answer to this question depends on the relative difference between the available technologies and on the intensity of the competition.

2 - Curated Contests

Sina Khorasani, Vanderbilt University, Nashville, TN, United States, Lakshminarayana Nittala, Vish Krishnan

To leverage the contest mechanism for sourcing quality solutions to complex problems, we propose a hybrid approach called the Curated Contest - in which firms benefit from competition among contestants but also engage in curation, by sifting and screening intermediate contestant submissions and potentially offering various forms of support for developing proof of concept and validating ideas. We provide guidance to solution-seeking firms on when and how to organize such curated contests in the presence of information asymmetries between solution seekers and skeptical contestants. We find that screening can stimulate additional effort from contestants. Firms can gain solver credibility by fine-tuning the reward size and other support schemes including late-stage compensation and proof of concept cost-sharing. We present a managerial framework to help design a carefully curated contest, which can help firms realize more effort from contestants.

3 -

Innovating Under Pressure: a New Model for Understanding R&D Competition

Linsheng Zhuang, National University of Singapore Institute of Operations Research and Analytics, Singapore, Singapore, Jussi Keppo, Zhi Chen

In many technological competitions, a firm must develop more superior technology over its rivals to win the entire market. Otherwise, firms simply split the market. We leverage the tug-of-war contest model to study this phenomenon. We show that the equilibrium can be substantially different from the corresponding equilibrium of classical rank-based contests. We discuss the implications of our results in the context of technological competitions.

SB27

Summit - 347

Online Reviews and Performance

Invited Session

New Product Development

Chair: Morvarid Rahmani, Georgia Institute of Technology, Atlanta, GA, United States

1 - The Bright Side of Lower Quality: Evidence from Restaurant Exploration

Manuel Sosa, INSEAD, Fontainebleau, France, Clara Carrera, Victor Martínez-de-Albéniz

The value derived from hedonic goods is affected by reference effects at the time of consumption, usually in the form of quality standards. Consumption typically involves two steps: first, the consumer chooses a given good, among a pool of available choices; then, the consumer experiences the good and derives a satisfaction from it. Between both steps, consumers might build expectations about the good that might affect the ultimate realized utility. We investigate the role of quality references in this two- stage (choice-outcome) process. We develop a flexible framework for estimating quality references and their effect in choice and outcome, that can include consumers' own past experiences as well as that of others, and can give salience to more recent or more distant past experiences. Using novel longitudinal data from online restaurant reviews, we find evidence of quality loss aversion in the choice decision stage, in accordance with prospect theory. However, in the outcome stage, we do find evidence of the opposite to loss aversion, i.e., satisfaction is affected much less than one would expect when going to a lower quality restaurant. This is consistent with consumers adjusting their expectations downwards and suggests that expectation adjustment protects consumers when they experience a good of lower-than- reference quality. Our results challenge the implicit assumption made by most recommendation systems that the expectation building process after making a choice does not change the outcome, and imply that it may be better to patronize activities by alternating between high- and low-quality choices.

2 - Work-Life Balance Satisfaction and Employee Turnover: A Large-Scale Analysis

Hyunwoo Park, Seoul National University, Seoul, Korea, Republic of, Morvarid Rahmani, Abhishek Deshmane

For firms to be able to compete in the complex and dynamic nature of today's economy, they need to continually innovate. Thereupon, it is important to understand the drivers of innovation at firms. In this paper, we study the effect of employee satisfaction on firms' innovation performance. We examine a social media platform, where employees voluntarily and anonymously share their outlook and opinions on various dimensions of their workplace environment. Using a sample of 905,930 employee reviews and 1,620,229 granted patents of 698 firms, we find that employee satisfaction with their workplace environment can significantly impact firms' innovation performance. More specifically, we find that while employee satisfaction with career opportunities enhances firms' innovation performance, their satisfaction with work/life balance negatively affects firms' innovation performance. Further, we find that employee satisfaction with their compensation and benefits can significantly reduce the value of firms' innovation in terms of both scientific and economic values, but it does not significantly affect the quantity of innovation outcomes. We also conduct analysis on sub-sample of firms based on their characteristics (ie, revenue, geographical location, and industry), and discuss the implications of our findings for research and practice.

3 - Timing the Tide: How the Chronology of External Conflicts Influences the Trajectory of Collaborative Projects (Tbd)

Shizhen Chen, Emory University, Atlanta, GA, United States, Tian Chan, Anandhi Bharadwaj, Gerald Kane

Building on the intriguing dynamics within online communities (OCs) and their ability to mobilize for collaborative action, our study extends the exploration into how external conflicts or crises affect the success trajectory of collaborative projects, especially when considering the timing of such disruptions. Through a comprehensive analysis, we examine the dual-edged nature of external crises, which can either hinder or enhance project outcomes based on their occurrence at different project stages. Drawing on various sources related to Reddit's r/place 2022 event, we delve into the initial devastating impact of early-stage conflicts that can jeopardize project viability, juxtaposed with the late-stage conflicts that can invigorate participation and expedite project success. Our findings highlight the critical role of conflict timing, offering a nuanced understanding of how external pressures can serve as a catalyst for engagement and innovation in collaborative endeavors. This research not only contributes to the discourse on conflict management within project management literature but also offers practical insights for navigating the complexities of collaborative project dynamics in the face of external challenges.

4 - How Women Promote Greater Social Responsibility on Social Media

Kejia Hu, Oxford University, Oxford, United Kingdom, Huibin Du, Lu Kong, Xiang Li

Problem definition: Social media is a platform for equal expression, with women significantly influencing social engagement. However, little research explores whether women's voices on social media lead to greater social responsibility contributions.

Academic/Practical relevance: This research expands the Individual Social Responsibility (ISR) discourse by highlighting ordinary women's role in promoting social responsibility through social media. The findings offer strategic insights for Corporate Social Responsibility (CSR), crisis management, and customer relationship management (CRM).

Methodology: Using two datasets from Sina Weibo, the study employs propensity score matching (PSM) and regression models to analyze gender differences in posting behaviors that promote social responsibility. The primary dataset consists of 225,974 original posts related to 26 public emergencies from 2017 to 2020, while the second captures user characteristics from 10 million historical posts of 138,282 users.

Results: Women post about public emergencies on social media more promptly than men, with posts appearing an average of 29 minutes earlier. Women's posts generate higher engagement, receiving 24.2% more reposts and 18.8% more comments than men's. Higher engagement is due to gender homophily, women's central positions in social networks, and their tendency to create original content, enhancing their effectiveness in promoting social responsibility.

Managerial implications: Firms should involve female employees in public relations and social media teams to enhance communication strategies. Social media platforms and policymakers should refine algorithms and policies to ensure equitable representation of women's voices, especially in discussions on social responsibility, optimizing organizational responses during emergencies.

SB28

Summit - 348

Sustainable Sourcing, Consumer Response, and Environmental Risks

Invited Session

MSOM: Sustainable Operations

Chair: Milind Sohoni, University at Buffalo, The State University of New York, Buffalo, NY, 14260, United States

Co-Chair: Abhinav Shubham, Georgia Institute of Technology, Atlanta, GA, United States

1 - Should I Curate or Not: A Game-theoretic Analysis of P2P Resale Marketplaces

Aditya Vedantam, State University of New York at Buffalo, Williamsville, NY, United States, Emre Demirezen, Subodha Kumar

The second-hand or resale market (ranging from automotive, luxury watches, clothing, and furniture among others) is growing rapidly. However, the growth of online marketplaces has also come with concerns around increasing amount of low quality and defective products being sold over resale marketplaces. In this paper we contrast two strategies adopted by peer-to-peer (P2P) in third party resale marketplaces: to curate used products by quality (e.g., thredUP), or not curate by quality (e.g., Tradesy, Poshmark). We investigate profitability and social welfare implications of quality curation and provide prescriptions for policymakers debating about regulating low quality and defective items on third party marketplaces.

2 - Sustainability Label: Product Quality Uncertainty and Post-Purchase Behaviors

Hyunsuk Baek, Arizona State University, Tempe, AZ, United States

We study how the quality uncertainty of a CSR-labeled product affects consumer's return and future purchase intention for another CSR-labeled product. We develop an analytical model which describes consumer's three sequential choices, with psychological factors, CSR belief and moral credit. We conduct a sequential discrete choice experiment while measuring participants' CSR belief and moral credit, in order to provide verified insight.

3 - Does Blockchain Facilitate Responsible Sourcing? An Application to the Diamond Supply Chain

Dmitrii Sumkin, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, United States, Sameer Hasija, Serguei Netessine

Blockchain technology has become widely accepted to demonstrate the provenance of physical goods, but there are open questions about its practical implementation and overall effect on ethical sourcing. In the diamond industry, blockchain enables credibility of the certificate of origin and therefore, allows to charge a premium for responsibly produced goods. Thus, one would expect blockchain to be the enabler of responsible sourcing. Using a stylized economic model, we demonstrate that, counter-intuitively, blockchain implementation may reduce incentives for customers to resell diamonds on the secondary markets and increase the cost of market segmentation. As a result, blockchain implementation could change a retailer's market segmentation strategy by increasing incentives to source from non-responsible suppliers. We also demonstrate that to reduce unintended consequences of blockchain implementation, the social planner should offer blockchain as an option (and not as a requirement) to a retailer. This work shows that caution is needed with wide-scale blockchain implementation. Although it is commonly recognized that blockchain's major application is in enhancing the traceability of durable goods or goods that should have a responsible supplier, we show that, in contrast, when factors of durability and responsibility come together, blockchain may negatively influence the responsibility level.

4 - Playing Fair? Environmental Impacts and Practices of Facilities in Minority Communities

Abhinav Shubham, Georgia Institute of Technology, Atlanta, GA, United States, Ravi Subramanian

Drawing on comprehensive US EPA and Census data, we examine the association between presence of substantial racial minority populations in host communities and facility-level environmental impacts and impact-reduction strategies. Our findings offer evidence for firms and policy makers to consider fairness and equity in managing and regulating environmental risks.

5 - Sustainability ratings in the supply chain

Junfei Lei, INSEAD, FONTAINEBLEAU, France

We study a novel business model in which a rating firm provides sustainability ratings for suppliers upon request of the buyers to minimize business risk under future regulation. The rating firm charges the suppliers and the buyers for the rating and endogenously determines what qualifies a supplier to be sustainable. The suppliers and buyers self-select whether to join the rating firm system. We identify conditions under which the rating firm drives a sustainable performance (by setting a stringent sustainability metric or by helping suppliers improve) and how system parameters (i.e., inherent risk, improvement costs, the magnitude of buyer-supplier transaction, and underlying uncertainty in sustainability performance) affect the same. We provide recommendations for the rating firm, capital markets, and policy.

SB29

Summit - 420

Optimization Society's Award Session I

Award Session

Optimization

Chair: Oktay Gunluk, Georgia Tech, Atlanta, GA, United States

1 - Wasserstein Distributionally Robust Optimization with Heterogeneous Data Sources

Daniel Kuhn, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland

We study decision problems under uncertainty, where the decision-maker has access to K data sources that carry biased information about the underlying risk factors. The biases are measured by the mismatch between the risk factor distribution and the K data-generating distributions with respect to an optimal transport (OT) distance. In this situation the decision-maker can exploit the information contained in the biased samples by solving a distributionally robust optimization (DRO) problem, where the ambiguity set is defined as the intersection of K OT neighborhoods, each of which is centered at the empirical distribution on the samples generated by a biased data source. We show that if the decision-maker has a prior belief about the biases, then the out-of-sample performance of the DRO solution can improve with K—irrespective of the magnitude of the biases. We also show that, under standard convexity assumptions, the proposed DRO problem is computationally tractable if either K or the dimension of the risk factors is kept constant.

2 - Fantastic Lifts and Where to Find Them: Using Smooth Parametrizations for Nonsmooth Optimization

Eitan Levin, Caltech, Pasadena, CA, United States

Many optimization problems are reformulated via smooth parametrizations, or lifts, of their domains. Examples include optimization over low rank matrices and tensors by optimizing over a factorization, the Burer--Monteiro approach to low rank semidefinite programs, and training neural networks by optimizing over their weights and biases. Such lifts give rise to two optimization problems, one optimizing a smooth cost function over a nonsmooth domain, and the other minimizing the composition of that cost with a lift map over a smooth manifold. The latter problem is smooth, and can be tackled with standard algorithms which may find various desirable points, like local minima or stationary points. In this talk, we ask when do these desirable points for the smooth problem map to desirable points for the nonsmooth problem. Remarkably, we show that the relationships between the two sets of desirable points is often independent of the cost function, and is rather a geometric property of the lift map itself. We therefore introduce a framework to study the effects of lifts on landscapes, and discuss its implications for the above examples.

3 - Asymptotic Normality and Optimality in Nonsmooth Stochastic Optimization

Liwei Jiang, Georgia Institute of Technology, Atlanta, GA, United States

In their seminal work, Polyak and Juditsky showed that stochastic gradient descent (with averaging) for minimizing smooth strongly convex functions enjoys a central limit theorem. Moreover, it has since been argued that the asymptotic covariance of the method is best possible among any estimation procedure in a local minimax sense of Hájek and Le Cam. A long-standing open question in this line of work is whether similar guarantees hold for important constrained/non-smooth problems, such as stochastic nonlinear programming. In this talk, we show that this is indeed the case.

SB30

Summit - 421

Network Flows and Parametric Cuts in Machine Learning and Advanced Optimization

Invited Session

OPT: Computational Optimization and Software

Chair: Dorit Hochbaum, UC Berkeley, Berkeley, CA, United States

1 - A Fast and Effective Breakpoints Algorithm for the Quadratic Knapsack Problem

Dorit Hochbaum, UC Berkeley, Berkeley, CA, United States

The maximum Quadratic knapsack problem (QKP) is to select from a given set of elements a subset of total cost bounded by a given budget, so that the sum of pairwise utilities of the pairs of selected elements and singleton utilities is maximized. We introduce here the Breakpoints algorithm for QKP, QKBP, that generates the concave piecewise linear envelope of the solutions to the relaxation of the problem for all values of the budget. The breakpoints in the concave envelopes are provably optimal solutions for the respective budgets, and for consecutive breakpoints the respective solutions are nested. For a given budget that falls between breakpoints we apply a simple greedy procedure that generates fast and high quality solutions. The breakpoints algorithm is highly scalable since the breakpoints and the concave envelope are attained using a parametric cut procedure that is very efficient. The performance of the parametric cut is further enhanced by a compact formulation of the problem. An extensive computational study presented here demonstrates that QKBP delivers consistently high quality solutions on known collections of benchmark instances, with much faster run times than state-of-the-art techniques. For newly introduced collections and large scale benchmark instances, the breakpoints algorithm delivers solutions that other state-of-the-art methods are either incapable of solving, or deliver poor quality even when provided with high time limits. All the benchmarks, algorithms used and results are publicly available on Github.

2 - Automatic Algorithm Selection for Pseudo-Boolean Optimization with Given Computational TIME Limits

Catalina Pezo Vergara, University of California, Berkeley, Berkeley, CA, United States, Dorit Hochbaum, Julio Godoy, Roberto Asín-Achá

Machine learning techniques have been proposed to automatically select the best solver from a portfolio of solvers, for a given problem. These techniques are used to implement meta-solvers that receive, as input, the instance of a problem, predict the best-performing solver in the portfolio, and execute it to deliver a solution. Typically, the quality of the solution improves with a longer computational time. This has led to the development of anytime meta-solvers, which consider both the instance and a user-prescribed computational time limit. Anytime meta-solvers predict the best-performing solver within the specified time limit.

This presentation will focus on the design of anytime meta-solvers for the NP-hard optimization problem of Pseudo-Boolean Optimization (PBO), which generalizes Satisfiability and Maximum Satisfiability problems. We demonstrate via extensive empirical study that our anytime meta-solver, named PBO_MS, improves dramatically on the performance of Mixed Integer Programming solver Gurobi, which is the best-performing single solver in the portfolio. We then generalize the anytime meta-solver by predicting a given number $p \geq 1$ of best solvers in the portfolio and run these sequentially, each with equal share of the specified time limit. The outcome is the best (minimum) objective function value obtained by any of the predicted solvers. This anytime p-meta-solver is shown here to outperform both the anytime 1-meta-solver as well as a fixed selection of solvers by a wide margin.

3 - Positive-Unlabaled Learning with Pairwise Similarity and Parametric Minimum Cut

Torpong Nitayanont, UC Berkeley, Berkeley, CA, United States, Dorit Hochbaum

Positive-unlabeled (PU) learning is a binary classification problem in which the labeled set only consists of samples from the positive class. The unlabeled set is a mixture of samples from positive and negative classes. In this work, we solve the PU learning problem using a network flow-based method called Hochbaum's Normalized Cut (HNC). HNC partitions samples, both labeled and unlabeled, into two groups to achieve high intra-similarity within groups and low inter-similarity between the two, with a tradeoff parameter that balances the two objectives. The problem is solved as a parametric minimum cut on an associated graph representation, producing different optimal partitions for different tradeoff parameters. Among all the generated cuts, we select one of them based on the prior information on the positive percentage, which is provided to most PU learning methods. Our proposed method works even when we have access to a limited number of labeled samples, and when the labeled set comes from only one class. When needed, reliable negative samples can be extracted from the unlabeled set. Our experimental study shows that HNC achieves competitive results compared to other state-of-the-art PU learning methods.

4 - The Max-Cut Decision Tree: Improving on The Accuracy and Running TIME of Decision Trees and Random Forests

Jonathan Bodine, UC Berkeley, Berkeley, CA, United States, Dorit Hochbaum

The Max-Cut decision tree involves novel modifications to a standard decision tree implementation, which is a widely used method of classification. It is implemented with the novel Maximum Cut splitting metric, maximally separating observations of different classes along locally meaningful representations of the input features. Experiments show these modifications can dramatically improve classification accuracy, while significantly decreasing computational time. The Max-Cut decision tree is expected to dramatically advance decision trees, with experimental extensions to random forests.

SB31

Summit - 422

Machine Learning for Optimization - II

Invited Session

OPT: Machine Learning

Chair: Gonzalo Constante Flores, Purdue University, West Lafayette, IN, United States

1 - From Inverse Optimization to Feasibility to Erm

Sharan Vaswani, Simon Fraser University, Vancouver, BC, Canada, Saurabh Mishra, Anant Raj

Inverse optimization involves inferring unknown parameters of an optimization problem from known solutions and is widely used in fields such as transportation, power systems, and healthcare. We study the contextual inverse optimization setting that utilizes additional contextual information to better predict the unknown problem parameters. We focus on contextual inverse linear programming (CILP), addressing the challenges posed by the non-differentiable nature of LPs. For a linear prediction model, we reduce CILP to a convex feasibility problem allowing the use of standard algorithms such as alternating projections. The resulting algorithm for CILP is equipped with a linear convergence guarantee without additional assumptions such as degeneracy or interpolation. Next, we reduce CILP to empirical risk minimization (ERM) on a smooth, convex loss that satisfies the Polyak-Lojasiewicz condition. This reduction enables the use of scalable first-order optimization methods to solve large non-convex problems while maintaining theoretical guarantees in the convex setting. Finally, we experimentally validate our approach on both synthetic and real-world problems and demonstrate improved performance compared to existing methods.

2 - Towards Practical Physics-Informed ML Design and Evaluation for Power Grid

Amritanshu Pandey, University of Vermont, Burlington, VT, United States, Shimiao Li, Larry Pileggi

When applied to a real-world safety critical system like the power grid, general machine learning methods suffer from expensive training, non-physical solutions, and limited interpretability. To address these challenges for power grids, many recent works have explored the inclusion of grid physics (i.e., domain expertise) into their method design, primarily through including system constraints and technical limits, reducing search space and defining meaningful features in latent space. Yet, there is no general methodology to evaluate the practicality of these approaches in power grid tasks, and limitations exist regarding scalability, generalization, interpretability, etc. This work formalizes a new concept of physical interpretability which assesses how a ML model makes predictions in a physically meaningful way and introduces an evaluation methodology that identifies a set of attributes that a practical method should satisfy. Inspired by the evaluation attributes, the paper further develops a novel contingency analysis warm starter for MadIoT cyberattack, based on a conditional Gaussian random field. This method serves as an instance of an ML model that can incorporate diverse domain knowledge and improve on these identified attributes. Experiments validate that the warm starter significantly boosts the efficiency of contingency analysis for MadIoT attack even with shallow NN architectures.

3 - Self-Supervised Learning for Constrained Optimization with Hard Linear Constraints

Hao Chen, Purdue University, West Lafayette, IN, United States, Gonzalo Constante Flores, Can Li

Over recent years, the success of machine learning has sparked a growing interest in leveraging neural networks to efficiently predict near-optimality yet feasible (or close-to-feasible) solutions. However, it is challenging to generate datasets of large optimization instances and strictly guarantee the feasibility of model inference. In this talk, we present a novel self-supervised and hard-constrained learning method by incorporating the proximal gradient descent algorithm into the deep learning framework. This approach ensures hard constraints by alternating two types of projection layers in the neural network, which always converges to a feasible point. This eliminates the need for optimal solutions of large optimization instances in the loss function and allows for self-supervised learning. The proposed model is demonstrated in some benchmarks, such as DC optimal power flow (DCOPF), exhibiting excellent performance in terms of optimality gap and solution feasibility. Model inferences on any incoming instance strictly satisfy linear constraints through projections with controllable violations and can be provided instantly through GPUs. Our numerical experiments indicate that these approximations are remarkably close to the optimal solutions, suggesting that the approach offers a reliable and efficient way to address large but time-sensitive linearly constrained optimization problems.

4 - Fully First-Order Methods for Constrained Bilevel Optimization

Kai Wang, Georgia Institute of Technology, Atlanta, GA, United States, Guy Kornowski, Swati Padmanabhan, Zhe Zhang, Suvrit Sra

Algorithms for bilevel optimization often encounter Hessian computations, which are prohibitive in high dimensions. While recent works offer first-order methods for unconstrained bilevel problems, constrained bilevel optimization remains relatively underexplored. We present the first fully first-order constrained optimization methods with finite-time hypergradient stationarity guarantees. For linear equality constraints, our algorithm converges to an ε-stationary point in $\widetilde{O}(ε-2)$ gradient oracle calls, which is nearly-optimal. For general convex inequality constraints, we attain (δ,ε)-Goldstein stationarity in either $\widetilde{O}(δ-1ε-4)$ or $\widetilde{O}(d{δ-1ε-3)$ gradient oracle calls, where d is the upper-level dimension. Along the way, we develop novel nonsmooth nonconvex optimization methods with inexact oracles. Our preliminary numerical experiments verify these theoretical convergence guarantees.

SB32

Summit - 423

Recent Advances in Data-driven Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Siva Ramani, Rice University, Houston, TX, United States

1 - Robustness Versus Statistical Efficiency: Superiority of Naive Optimization via Worst-Case Stochastic Dominance

Arindam Roy Chowdhury, Columbia University, New York City, NY, United States, Henry Lam

Many decision-making tasks give rise to optimization problems with uncertainty in the underlying parameter or probability distribution. This uncertainty can arise from statistical noises due to limited data, or non-stationary parameter shifts that are not predictable from the past. In the literature, various approaches ranging from (distributionally) robust optimization to regularization have been proposed, as attempts to obtain solutions that perform well statistically or robustly against model shifts. Contrary to common belief, we show that, in terms of the behavior of regret or optimality gap, a naive optimization formulation based on empirical or simple point estimate of the uncertain parameter is in a sense optimal, regardless of whether the uncertainty comes from statistical noises, unpredictable shifts, or both. Our superiority assertion for naive optimization is based on a new notion of worst-case stochastic dominance for the asymptotic regret, and is argued via what we call the "symmetry" of uncertainty that applies universally over statistical errors, adversarial errors, and their interactions. We show how our results apply to the comparisons among a wide range of existing data-driven and robust optimization formulations.

2 - Approximation Algorithms for Distributionally-Robust Combinatorial Optimization

Chaitanya Swamy, University of Waterloo, Bangalore, ON, Canada, Andre Linhares

In two-stage stochastic optimization, we have a probability distribution over possible realizations of the data, and we seek to take decisions both before and after a scenario is realized to typically minimize the total expected cost. We consider the popular distributionally-robust paradigm: we are now given a collection \C of distributions, and the goal is to minimize the maximum expected total cost with respect to a distribution in \C. There has been almost no prior work on developing approximation algorithms for distributionally-robust combinatorial-optimization problems, wherein the scenario-set is discrete and the underlying deterministic problem is often NP-hard. We provide a framework for designing approximation algorithms in such settings when the collection \C is a ball around a central distribution accessible only via a sampling black box. We devise

algorithms that draw polynomially many samples and return solutions with provable worst-case quality guarantees. We show that an LP-relaxation of the problem can be reduced to a standard 2-stage problem, which can then be (approximately) solved given an (approximation) algorithm for a certain max-min problem akin to the problem of finding the worst-case scenario in robust optimization. Complementing this with LP-rounding ideas, we obtain the first approximation algorithms for various distributionally-robust combinatorial-optimization problems including set cover, facility location, Steiner tree. Our guarantees are often within O(1)-factors of the guarantees known even for the deterministic version of the problem.

3 - A Unified Framework to Obtain Data-Driven Performance Guarantees for Robust Markov Decision Processes

Siva Ramani, University of Washington, Seattle, WA, United States

We introduce a family of distance-based s-rectangular robust MDPs (s-RMDPs), where in each state, the ambiguity set equals a sublevel set of the norm of a vector of distances from reference pmfs. The construction ensures that the distance-based (s,a)-rectangular robust MDP ((s,a)-RMDP) is contained within this family. In the first part, we rigorously analyze the relative conservativeness between (s,a)- and s-RMDPs in the non data-driven context. Next, we study the data-driven versions of our s-RMDPs. Under a unifying framework, we prove that our s-RMDPs satisfy all the data-driven performance guarantees that are deemed essential for any data-driven optimization framework. Moreover, this framework enables us to make sharp comparisons between (s,a)- and s-RMDP in the data-driven setup. An artifact of the analyses behind these guarantees is the surprising conclusion that (s,a)-RMDPs might be the least conservative among all s-RMDPs within our family. Finally, we compare our data-driven robust framework with other data-driven approaches for MDPs.

4 - Multistage Chance-Constrained Programming

Soumya Ranjan Pathy, Clemson University, Clemson, SC, United States, Hamed Rahimian

In this talk, we study multistage chance-constrained programming, a critical area in optimizing decision-making under uncertainty. Our analysis considers two possibilities for enforcing chance constraints in the model: (i) constraints that are statically enforced only at the end of the planning horizon and (ii) constraints that are dynamically enforced throughout the planning horizon. We introduce novel decomposition-based computational approaches for solving such complex problems. Using a stylized inventory management problem, we demonstrate preliminary computational results that highlight the efficacy and potential of our proposed methods. Furthermore, we provide a comparative analysis of the outcomes derived from a multistage policy against those obtained from a conventional two-stage multiperiod model. This discussion emphasizes the added value and strategic advantage of implementing a multistage approach in real-world applications.

5 - Light Pareto Robust Optimization for Radiotherapy Treatment Planning

Houra Mahmoudzadeh, University of Waterloo, Waterloo, ON, Canada, Fahimeh Rahimi, Danielle Ripsman, Hossein Abouee Mehrizi

Robust optimization (RO) is well-known for immunizing operators against worst-case uncertainty realizations. The RO methodology, however, neglects non-worst-case scenarios, potentially leading to overly-conservative solutions. Pareto robust optimization (PRO) further optimizes for non-worst-case scenarios, generating robust solutions that may be less conservative for non-worst-case uncertainty realizations. In this paper, we extend the theory of PRO to practical settings where expert knowledge on likely scenarios exists. We define and characterize superior PRO solutions and develop algorithms that can partition the set of RO solutions to determine the Pareto robust frontier of superior PRO solutions. To further reduce the level of conservatism of the PRO solutions, we also introduce the concept of light PRO solutions, which sacrifice some worst-case performance to improve non-worst-case behavior while maintaining robustness guarantees. We quantify the gain in non-worst-case performance as a function of worst-case loss, discuss the theoretical properties of this gain function, and analyze the trade-off between worst-case optimality and non-worst-case performance.
We apply these methods to a radiation therapy planning problem for breast cancer patients. The proposed approach is shown to significantly improve the quality of treatment plans in non-worst-case scenarios while maintaining robustness and limiting the optimality sacrificed in the worst-case outcomes.

SB33

Summit - 424

Advances in Stochastic Programming: Theory and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Hanbin Yang, Chinese Universtiy of Hong Kong, Shenzhen, Shenzhen, N/A

1 - Enhanced Lagrangian Cuts for Multi-Stage Stochastic Mixed-Integer Programs

Hanbin Yang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Haoxiang Yang

This work addresses multi-stage stochastic mixed-integer linear programming (MS-MILP) with general mixed-integer state variables. We demonstrate that while vanilla Lagrangian cuts can approximate the convex envelope of value functions, they can be steep and fail to define a facet of this envelope. To overcome this limitation, we introduce enhanced Lagrangian cuts and an innovative iterative lifting/branching procedure. These enhanced cuts are obtained through a geometric approach linking the epigraph of value functions to cutting-plane methods, enabling the generation of tight Pareto-optimal cuts. Consequently, our method provides a superior lower approximation with fewer cuts. Moreover, we propose a convergent decomposition algorithm incorporating these enhanced cuts, making feasible the challenging task of solving general MS-MILP. Our computational results reveal the efficacy of the enhancement techniques: the proposed algorithm consistently achieves a small optimality gap within an acceptable time frame, marking a significant advancement in tackling MS-MILP problems.

2 - Equilibrium Existence and Algorithm of Discounted Non-Zero-Sum Stochastic Game Between Risk-Averse Players

Yufeng Gao, the University of Hong Kong, Hong Kong, Hong Kong

This paper presents a model for n-risk-averse-player finite state/action stochastic games. We consider non-zero-sum discounted stochastic games with uncertainty from both supply and demand, where players have risk-averse attitudes towards uncertainty. The measure of risk is conditional value-at-risk (CVaR). We prove the existence of equilibrium without complexity assumptions, and propose an algorithm to calculate equilibrium points. The algorithm is compared with traditional best response algorithm over multi-dimension performance.

3 - Dynamic Closed-Loop Supply Chain Network Design and Operations for Electric Vehicle Battery

Wentao Zhao, University of Southern California, Los Angeles, CA, United States, Maged Dessouky, Randolph Hall

Large-scale adoption of electric vehicles (EVs) can reduce greenhouse gas emissions and improve air quality. Nonetheless, a significant amount of EV batteries are expected to retire within five years, posing environmental and economic challenges regarding management and disposal. Implementing a closed-loop supply chain (CLSC) network is a promising strategy. The core idea is to enhance battery utilization and reuse valuable resources. In this work, we develop a bi-level computational framework for the dynamic CLSC network design and operations problem where two levels of decisions in different time resolutions are periodically optimized. At the upper level, we formulate a stochastic dynamic programming model to design a capacity expansion and facility-building policy for the CLSC to cope with the uncertainties from markets and politics. At the lower level, we formulate a stochastic multi-stage stochastic programming model to optimize the daily operations of the CLSC and maximize the overall revenue under demand and supply uncertainties. A heuristic approach is proposed to solve large-scale problems. The proposed heuristic is a local search algorithm embedded with sample average approximation. Lastly, we use real-world datasets to validate the model and demonstrate the solution approach.

4 - Multiscale Stochastic Programming with Applications on Power Systems

Yihang Zhang, University of Southern California, Los Angeles, CA, United States, Suvrajeet Sen

We introduce a multiscale stochastic programming framework tailored for power systems with high renewable penetration. This framework addresses the multiscale uncertainties inherent in renewable energy sources. Our approach employs multiscale modeling techniques to capture the dynamics of conventional and renewable generators as well as demand patterns across various scales. This enables us to form detailed operational strategies for short-term periods while maintaining a broader perspective for long-term planning. To coordinate decisions across different timescales in a manageable way, we introduce a context mechanism similar to the state concept in dynamic programming. Additionally, we propose a practical scenario-based decomposition method for efficient computational implementation.

SB34

Summit - 425

Recent Advances in Min-Max Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Mert Gurbuzbalaban, Rutgers University, Piscataway, NJ, United States

Co-Chair: Necdet Aybat, Penn State University, University Park, PA, United States

1 - AGDA+: Proximal Alternating Gradient Descent Ascent Method with Backtracking for Double-regularized Nonconvex-Strongly Concave Minimax Problems

N. Serhat Aybat, Penn State University, University Park, PA, United States

We consider double-regularized weakly convex-strongly concave (WCSC) minimax problems of the form (P) : min_x max_y g(x) + f (x, y) − h(y), where g, h are closed convex, f is L-smooth in (x, y) and (strongly) concave in y. We propose a proximal alternating gradient descent ascent method AGDA+ that can adaptively choose primal-dual stepsizes to compute an approximate stationary point for (P) without requiring the knowledge of the global Lipschitz constant L. This algorithm stands out by its ability to exploit the local Lipschitz structure and eliminates the need for precise tuning of hyper-parameters. AGDA+ achieves the optimal iteration complexity of O(eps^−2) with the best O(1) constant among the existing methods agnostic to L. The numerical experiments demonstrate its robustness and efficiency in both deterministic and stochastic scenarios. To the best of our knowledge, this is the first single-loop backtracking method for solving WCSC minimax problems.

2 - High probability and risk-averse guarantees for stochastic saddle point problems

Mert Gurbuzbalaban, Rutgers University, Piscataway, NJ, United States

We first consider stochastic strongly-convex-strongly-concave (SCSC) saddle point (SP) problems which frequently arise in applications ranging from distributionally robust learning to game theory and fairness in machine learning. We focus on the recently developed stochastic accelerated primal-dual algorithm (SAPD), which admits optimal complexity in several settings as an accelerated algorithm. We provide high probability guarantees for convergence to a neighborhood of the saddle point that reflects accelerated convergence behavior. We also provide an analytical formula for the limiting covariance matrix of the iterates for a class of stochastic SCSC quadratic problems where the gradient noise is additive and Gaussian. This allows us to develop lower bounds for this class of quadratic problems which show that our analysis is tight in terms of the high probability bound dependency to the parameters. We also provide a risk-averse convergence analysis characterizing the ``Conditional Value at Risk", the ``Entropic Value at Risk", and the χ2-divergence of the distance to the saddle point, highlighting the trade-offs between the bias and the risk associated with an approximate solution obtained by terminating the algorithm at any iteration.

3 - A Stochastic GDA Method With Backtracking For Solving Nonconvex (Strongly) Concave Minimax Problems

Qiushui Xu, Penn State University, State College, PA, United States

We propose a stochastic GDA method with backtracking (SGDA-B) to solve nonconvex-strongly-concave minimax problems $\min_x \max_y \sum_{i=1}^N g_i(x_i)+f(x,y)-h(y)$, where $\{g_i\}_{i=1}^N$, h are closed convex, and f is L-smooth and μ-strongly concave in y. We assume that one only has an access to an unbiased stochastic oracle of ∇f with a finite variance. Without requiring to know L, SGDA-B using random block-coordinate updates can compute an ϵ-stationary point within $\mathcal{O}(L\kappa^3 \epsilon^{−4} log(1/p))$ stochastic oracle calls, with probability at least 1−p, where $kappa=L/\mu$. We provide numerical results for SGDA-B on a distributionally robust learning problem.

4 - High probability bounds for stochastic non-convex mini-max optimization

Yasa Syed, Rutgers University, Piscataway, NJ, United States

Stochastic min-max optimization is prevalent across various disciplines, from engineering to game theory and machine learning. Despite recent progress, satisfactory high-probability guarantees for non-convex mini-max problems still do not exist to our knowledge. We will provide first-time high-probability guarantees for NC-PL mini-max problems (where the objective is non-convex with respect to the primal variable and satisfies a Polyak-Łojasiewicz (PL) inequality in the dual variable) and illustrate the results on several numerical examples including distributionally robust learning.

SB35

Summit - 427

Learning in Transportation Network Modeling

Invited Session

TSL: Intelligent Transportation Systems

Chair: Zhichen Liu, University of Michigan, Ann Arbor, United States

Co-Chair: Minghui Wu, University of Michigan, Ann Arbor, MI, United States

1 - An Unsupervised Learning-Based Branch-and-Price With Neural-Cuts Framework to Solve Dial-A-Ride Problem

Xinwu Qian, Rice University, Houston, TX, United States

The study presents the Neural Branch-Price-and-Cut (NeuralBPC) method, an ML-based approach that optimizes the Column Generation (CG) method by introducing Neural Cuts. Utilizing a Graph Neural Network (GNN) with an unsupervised learning model and a probabilistic loss function, this method overcomes traditional sample collection bottlenecks in ML solutions for optimization problems, which requires to completely solve the NP-hard problems as labels. The NeuralBPC uses previously generated columns and passenger-conflicting graphs to forecast effective cutting planes for the Restricted Master Problem (RMP), effectively reducing the search space of the Pricing Sub-Problem (SP), generating high-quality solutions with fewer CG iterations. By exploiting the iterative nature of CG for sample collection, the algorithm dynamically produces the most effective Neural Cuts, demonstrated through case studies on variants of Dial-a-Ride Problems (DARPs), enhancing solution quality and minimizing computational times.

2 - Learning from Global Satellite Imagery and Local Sensor Data for Enhanced Dynamic Origin-Destination Demand Estimation

Jiachao Liu, Carnegie Mellon University, Pittsburgh, PA, United States

The utilization of high-resolution satellite imagery is an increasingly prominent trend in transportation science, offering the potential to enhance existing dynamic network models and address the limitations of traditional data sources. Satellite images provide comprehensive coverage and detailed insights into road networks, traffic patterns, and infrastructure at a global scale that was previously unachievable with data collected from local sensors. A primary challenge lies in developing a robust calibration framework (i.e., dynamic origin-destination demand estimation, DODE) for large networks that effectively integrates traffic state information (i.e., traffic density) obtained from satellite imagery with other data sources (i.e., traffic count, speed). This study proposes a computational-graph-based DODE framework leveraging multi-source data including traffic densities derived from satellite images and explores the benefits of incorporating satellite imagery into model calibration through numerical experiments using both synthetic and real-world data.

3 - Linear Complementarity Systems for the Morning Commute Problem with Ridesharing and Dynamic Pricing

Wei Gu, University of Southern California, Los Angeles, CA, United States

Since Vickrey (1969), the morning commute problem has been studied for more than fifty years. The emerging ridesharing services and relevant infrastructures, such as High-Occupancy Toll (HOT) lanes, provide more flexibility for travelers, more opportunities for sustainable transportation systems, and at the same time, more challenges for this classical problem. To capture traffic dynamics, we propose a modified bottleneck model that avoids time-delayed terms in computing travel times and maintains desired mathematical properties. Then we develop a general mathematical modeling framework for the morning commute problem with ridesharing, including travel modes, infrastructures, and businesses. Formulated as Linear Complementary Systems (LCS), the proposed model simultaneously captures travelers' departure time choice, lane choice between HOT lane and general purpose lane, as well as mode choice between ridesharing and solo driving. We provide rigorous analysis of solution existence for the LCS-based general modeling framework. To approximate the proposed continuous-time model, discrete-time model is generated using an implicit time discretization scheme, with the theoretical guarantee to converge back to the original continuous-time form. Analytical solutions for dynamic prices, including drivers' incomes, passengers' payments, and HOT lane toll charges, are derived to balance the various demands of travelers, company, and society. The proposed models and dynamic pricing strategies are validated in numerical examples. Results show that our proposed methodology simultaneously benefits travelers, company, and society toward urban sustainability through ridesharing: smaller travel costs, positive cash flow for ridesharing businesses, and better system performance.

4 - Designing Robust Transportation Networks with Imperfect User Equilibrium Predictions

Zhichen Liu, University of Michigan, ANN ARBOR, MI, United States, Yafeng Yin

Forecasting current user equilibria accurately is challenging, and predictions based on current data often fail to align with actual future user equilibria, which are benchmarks for long-term transportation network design. This study proposes a robust network design framework that accommodates potential deviations between predicted and actual future equilibria. The optimized design from this framework ensures performance on the actual user equilibria despite the nonconvexity of the design problem. Moreover, to adapt this framework for real-world-sized road networks, we propose a novel computational-graph-based solution algorithm. This algorithm enables the efficient parallel computation of uncertain user equilibrium constraints and facilitates scalability.

SB36

Summit - 428

Human-centric Warehouse and Order Fulfillment Operations

Invited Session

TSL: Facility Logistics

Chair: Junsu Kim, Inha University, Incheon, Korea, Republic of

Co-Chair: Junsu Kim, Inha University, Incheon, Korea, Republic of

1 - Can Collaborative Robotic Solutions Make Store Fulfillment Economically Viable?

Joyjit Bhowmick, Rensselaer Polytechnic Institute, Troy, NY, United States, Jennifer Pazour, Iman Dayarian, Sebastian Köhler, Gideon Arndt, Kai Furmans

Omnichannel services, such as buy-online-pickup-in-store, curbside pickup, and ship-from-store, have shifted the order-picking tasks previously completed by in-store customers doing their own shopping to the retailer's responsibility. To fulfill these orders, many retailers have deployed a store fulfillment strategy, where online orders are picked from brick-and-mortar retail store shelves. However, low profit margins, labor shortages, and increasing labor costs make it difficult for stores to cost effectively to fulfill online orders through store fulfillment. Motivated by this challenge, we explore two order-picking system designs where collaborative robotic solutions augment the picking process of dedicated employees. The first deploys autonomous mobile robots (AMRs) synchronizing with in-store customers to pick online items, where the availability of in-store customers is stochastic and in-store customers are compensated for their picking effort. In the second, a modular and moveable robotic platform with a picking arm is deployed in the store having the ability to perform order picking and replenishment. We create operational and economic models that are fed with empirical data for shopping behavior and actual online orders to illustrate that order-picking policies with such automated solutions are operationally and economically viable, whether the store purchases the robots upfront or follows a subscription based approach taking wages, benefits, automated solution investments, maintenance, customer compensation, and customer hassle into consideration.

2 - An experimental study of large language model for human interactive robots in a warehouse management

HanByul Ryu, Inha University, Incheon, Korea, Republic of, Daisik Nam

To address the increasing demand in logistics and ensure efficient logistics operations, warehouses are being automated. While logistics robots are being utilized as one of the strategies for automation, they lack flexibility compared to human workers, making efficient operations challenging. Consequently, human intervention is frequently required even within automated systems. This study proposes the utilization of Large Language Models (LLMs) as an intermediate solution to address collaboration between robots and workers and introduces its framework. LLMs like GPT utilize advanced artificial intelligence techniques to effectively interpret human instructions, offering the opportunity for flexible adaptation of robots within warehouse environments. Simultaneously, LLMs hold the potential to offer solutions for better warehouse operations such as save labors to control robots. This study discusses whether relying solely on LLMs for warehouse operations is feasible, or if a hybrid approach, integrating existing human control alongside tailored inputs for LLMs, would be more appropriate. The paper explores the potential benefits and challenges of each approach, aiming to provide insights into the most effective utilization of LLMs in warehouse management. This study can contribute to exploring collaborative approaches between humans and technology in warehouse operations before complete automation is achieved.

3 - The process of tailoring work schedules in warehouses considering the perceived physical fatigue levels

Junsu Kim, Inha University, Incheon, Korea, Republic of, Hosang Jung, SEUNGSIK OH

This study focuses on the process of tailoring work schedules for warehouse workers by incorporating their perceived physical fatigue levels. In light of the importance of understanding and mitigating workplace fatigue, we present a practical and systematic methodological framework for this process. This approach integrates survey-based fatigue measurements, regression analysis, and mathematical optimization modelling. A survey of order pickers in a South Korean order fulfillment center and the results of a corresponding regression analysis provide valuable insights regarding the process of tailoring work schedules to manage fatigue. Moreover, the optimization model, which takes into account realistic operational constraints, determines the duration of work cycles and the timing of work and break cycles. By using this regression model to inform the optimization model, we generate four distinct schedules that navigate between the goals of minimizing the average fatigue levels of workers and minimizing their final fatigue levels. The model's focus can be adjusted to suit practitioners' needs. After consulting with the center manager, we approved a schedule that focused on reducing workers' mean final fatigue levels. As compared to the current schedule, the selected new schedule reduced workers' final fatigue level by 11.5% when the same number of total hours were worked, thus highlighting the method's potential to enhance worker well-being.

4 - Optimization of Two-tier Robotized Sorting Systems for Parcel Delivery

Reem Khir, Purdue University, West Lafayette, IN, United States, Junsu Kim

This study addresses an operational planning challenge aimed at improving the parcel sortation process in two-tier robotized sorting systems, a forefront technology in e-commerce warehouse automation. The system's upper tier transports parcels from loading to drop-off points using autonomous mobile robots, while the lower tier organizes parcels into roll containers by their final destinations for loading onto outbound trucks. These processes are interconnected assignment problems, requiring holistic decision-making to maximize automation benefits and maintain productivity. We introduce a mixed-integer linear programming model to optimize operational coordination across the interdependent tiers, considering the incomplete information about parcel arrival sequences. By employing a tailored multiple-scenario approach, we develop the solution procedure to devise sortation plans that effectively manage the uncertainty of arrival sequences. The computational effectiveness of our approach remains robust even in larger-sized problem instances. Furthermore, our comprehensive analysis explores the value of information on incoming arrival sequences and the adaptability of assignment decisions, providing essential insights for managers at companies using or considering two-tier robotized sorting systems.

SB37

Summit - 429

Data-driven and Optimization Foundations for Emerging Mobility Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Devansh Jalota, Stanford University, Stanford, CA, United States

Co-Chair: Matthew Tsao, Lyft, San Francisco, CA, United States

Co-Chair: Gioele Zardini, Stanford University, Stanford, United States

1 - Sustainability in Online Freight Marketplaces

Saurabh Amin, MIT, Cambridge, MA, United States, Maxime Bouscary, Mazen Danaf

Road freight transport moves about 26.8 trillion tonne-kilometers annually, mainly in trucks and delivery vehicles, emitting around 2.2 billion tonnes of CO_2. The trucking sector faces the dual challenges of increasing demand from e-commerce and on-time delivery services and supply-side constraints due to rising fuel costs and fledgling infrastructure to support electric trucks. We investigate the potential role of digital brokerage platforms in reducing both operational costs and empty miles by strategic bundling of loads and dynamic pricing over a finite horizon. Our approach accounts for the heterogeneities in carriers' arrival locations and preferences over quality and price of load bundles. We provide new asymptotically optimal bounds on the maximal expected revenue in this general setting and leverage these bounds to tractably select load bundles. We identify operational regimes under which optimally priced load bundles lead to simultaneous improvement in both platform revenue and carrier satisfaction, while reducing empty miles.

2 - Learning and Control for Safety, Efficiency, and Resiliency of Embodied AI

Fei Miao, University of Connecticut, Storrs, CT, United States

With rapid evolution of sensing, communication, and computation, integrating learning and control presents significant Embodied AI opportunities. However, current decision-making frameworks lack comprehensive understanding of the tridirectional relationship among communication, learning and control, posing challenges for multi-agent systems in complex environments. In the first part of the talk, we focus on learning and control with communication capabilities. We design an uncertainty quantification method for collaborative perception in connected autonomous vehicles (CAVs). Our findings demonstrate that communication among multiple agents can enhance object detection accuracy and reduce uncertainty. Building upon this, we develop a safe and scalable deep multi-agent reinforcement learning (MARL) framework that leverages shared information among agents to improve system safety and efficiency. We validate the benefits of communication in MARL, particularly in the context of CAVs in challenging mixed traffic scenarios. We incentivize agents to communicate and coordinate with a novel reward reallocation scheme based on Shapley value for MARL. Additionally, we present our theoretical analysis of robust MARL methods under state uncertainties, such as uncertainty quantification in the perception modules or worst-case adversarial state perturbations. In the second part of the talk, we briefly outline our research contributions on robust MARL and data-driven robust optimization for sustainable mobility. We also highlight our research results concerning CPS security. Through our findings, we aim to advance Embodied AI and CPS for safety, efficiency, and resiliency in dynamic environments.

3 - Combinatorial Optimization Augmented Machine Learning for Contextual Multi-Stage Problems

Maximilian Schiffer, Technical University of Munich, Munich, Germany

Combinatorial optimization augmented machine learning (COAML) is a novel field that combines methods from machine learning and operations research to tackle contextual data-driven problems that involve both uncertainty and combinatorics. These problems arise frequently in industrial processes, where firms seek to leverage large and noisy data sets to optimize their operations. COAML typically involves embedding combinatorial optimization layers into neural networks and training them with decision-aware learning techniques. This talk provides an overview of the underlying paradigm, algorithmic pipelines, and foundations based on selected application cases. Particularly, I will demonstrate the effectiveness of COAML on contextual and dynamic stochastic optimization problems, as evidenced by its winning performance on the 2022 EUROMeetsNeurIPS dynamic vehicle routing challenge.

4 - Markov Potential Games in on-Demand Mobility Services

Kenan Zhang, EPFL, Lausanne, Switzerland

Markov game is a game theoretical framework to describe the multi-agent interaction in a Markovian dynamic system. As a special case of Markov game, Markov potential game (MPG) features special properties on the agent utility such that finding the stationary Nash equilibrium can be reduced to solving a global optimization problem. In this talk, I will present our recent work on MPG in the context of on-demand mobility services and discuss how this framework facilitates the analysis of service design and operations, including market equilibrium, dynamic pricing, relocation incentives, etc.

SB38

Summit - 430

Reliable and Scalable Urban Air Mobility

Invited Session

Aviation Applications

Co-Chair: Minghao Chen, Columbia University, 514 W 114 St, New York, 10025, United States

Co-Chair: Shulu Chen, George Washington University, Washington, DC, United States

1 - V-Hatt: An Integrated UAM Scheduling and Vertiport Management System

Shulu Chen, George Washington University, Washington, DC, United States

This paper presents the Vertiport Human Automation Teaming Toolbox (V-HATT), a novel framework for managing Urban Air Mobility (UAM) within terminal airspace and vertiports. V-HATT is designed to integrate the efforts of human vertiport operators with an automated system, enhancing scheduling and real-time operational control to establish a unified traffic management system. Utilizing the Helo Holdings, Inc. (HHI) Heliport as a case study, we demonstrate V-HATT's capability to effectively manage arrival and departure in densely populated urban airspaces such as New York City. The framework addresses all three critical phases of UAM operations: the planning phase focuses on the design of terminal airspace and vertiport layouts and on measuring airspace capacity; the scheduling phase employs mixed-integer linear programming (MILP) algorithms to optimize arrival and departure sequences and to efficiently manage vertiport ground resources on a rolling horizon, aiming to minimize delays and enhance the utilization of ground resources while considering specific constraints of vertiports and aircraft; the tactical operation phase supports the simulation of UAM operations in real-time and provides interfaces for human operators to issue maneuver advisories and adjust schedules to address unforeseen events. This integrative approach promises significant improvements in UAM efficiency, safety, and reliability, thereby contributing to the sustainable expansion of urban air transportation systems.

2 - A Hybrid Approach for in-Flight Li-Ion Battery Soc Prediction

Abenezer Taye, George Washington University, Washington DC, DC, United States

Lithium-ion (Li-ion) batteries are increasingly used to power unmanned aerial vehicles (UAVs) due to their high energy capacity and long lifespan. Accurately predicting the State of Charge (SoC) of these batteries, which estimates their remaining capacity as a percentage of total capacity, is essential for ensuring flight safety and effective planning of UAV operations. Traditional SoC prediction methods fall into two categories: model-based, which relies on physical models of the aircraft and battery but is computationally intensive and slow for in-flight use, and data-driven, which is faster but less accurate over longer prediction horizons. Our research introduces a hybrid method that combines the accuracy of model-based approaches with the efficiency of data-driven methods. By integrating a multi-rotor power consumption model with a Long Short-Term Memory (LSTM) based prediction scheme, our method uses projected power requirements and current battery state to predict SoC, allowing for dynamic flight mission planning. This approach was tested in simulated flight scenarios, demonstrating both high accuracy and computational efficiency compared to model-based methods, offering significant improvements in UAV safety and energy management.

3 - Flight Planning and Strategic Deconfliction Under Wind Hazards

Amin Tabrizian, The George Washington University, Washington, DC, United States

Winds pose challenges for small Unmanned Aircraft Systems (UAS) and electric vertical take-off and landing (eVTOL) aircraft, especially due to their lighter weight and lower altitude operations. Wind conditions can rapidly change in the lower atmosphere, affecting trajectories and increasing energy consumption. To address this, we develop flight planning algorithms to adjust routes and departure times, avoiding wind hazard areas.

We utilize Version 4.2 of the Weather Research and Forecasting model to predict wind conditions at a high resolution. These forecasts are processed using Kernel Density Estimation to identify regions where wind speeds exceed safety thresholds. Density-based clustering techniques help delineate unsafe airspace areas.

Our routing algorithm minimizes flight time while avoiding wind hazard polygons and operational volumes of other aircraft. Using the A* optimal cost path algorithm and fast marching tree algorithm, we generate efficient routes while ensuring deconfliction with other airspace users.

We apply our approach in the Dallas Fort Worth area, demonstrating its effectiveness for small UAS delivery and urban air mobility. Our findings highlight the algorithm's ability to suggest optimal ground delays and alternative routes, respect operational volumes, and mitigate system-wide delays and traffic pattern changes caused by diverse wind hazards.

4 - Energy-constrained, Risk-averse Drone Routing under Spatiotemporally Varying Urban Winds

Minghao Chen, Columbia University, New York, NY, United States, Max Li, Marco Giometto, Andrew Smyth

Urban areas present unique challenges for drone-based deliveries due to the spatially and temporally variable nature of the wind field at low altitudes. In this paper, we propose a novel approach to address these challenges and ensure safe and reliable drone delivery routings in windy urban environments. Our approach incorporates the stochastic and spatially heterogeneous urban wind field using scenario-based stochastic programming and probabilistic modeling techniques. We develop stochastic mixed integer linear programs (SMILPs) that integrate realistic wind scenarios generated via large-eddy simulations (LESs) with the stochastic drone routing problem (SDRP). We employ the parallel large neighborhood search algorithm, which effectively explores the solution space and handles hundreds of customers' requests in a short time. We compare the performance of heuristics implementation with SMILPs implementation on small cases and show its superiority in terms of solution quality. Through simulations and performance evaluations on large-scale instances, we demonstrate the effectiveness of our approach under different wind conditions and number of customers. We also conduct a sensitivity analysis to investigate the influence of operational altitude (i.e., the height of the wind field) on the performance of our method. Our results provide valuable insights for optimizing drone routing decisions in windy urban environments.

SB39

Summit - 431

2024 RAS Student Paper Competition

Invited Session

Railway Applications

Chair: Erick Wikum, Wikalytics LLC, 4906 Whispering Creek Ct, Maineville, OH, 45039, United States

Co-Chair: Nikola Markovic, University of Utah, Salt Lake City, UT, United States

1 - Rail Applications Section Student Paper Competition Finalists

Erick Wikum, Wikalytics, LLC, Maineville, OH, United States

The competition rules, number of entries, judging, and types of papers that were submitted will be reviewed

2 - Combining Mixed Integer Programming with Row Generation for Railway Network Capacity Assessment

Christopher Szymula, Technische Universität Dresden, Dresden, Germany, Nikola Besinovic, Karl Nachtigall

Railway network capacity assessment for large-scale networks with explicit regard of the intended operations proves crucial for strategic network planning and development. We present the MIP-based railway network utilization model (MIP-RNUM), which assesses the network's practical capacity with respect to given service characteristics. It is based on Petri Nets (PN) and combines the benefits of mixed-integer linear programming and PN-modelling. Based on the structures of critical process cycles,the presented MIP-RNUM is then used to assess the network's capacity. For better scalability on real-life instances, we extend it to the MIP-RNUM-RG, arow generation solution approach which considers the decisive set of critical process cycles only.The feasibility of every solution during the row-generation procedure is monitored by a newly introduced max-plus based lower bound. Our model computes the specific network'scapacity,regarding for the corresponding operational scenario. The results show the applicability of the MIP-RNUM-RG for network capacity assessment of real-life instances. As such, the model provides a comprehensive, applicable, and extendable solution to the capacity assessment of large and complex railway networks.It thus contributes to the recent efforts to strengthen the railway mode's share on the modal split of European transportation.

3 - Joint Rolling Stock and Crew Scheduling in Urban Rail Networks

Entai Wang, Beijing Jiaotong University, Beijing, China, People's Republic of, Lixing Yang, Yossiri Adulyasak, Jean-François Cordeau, Ziyou Gao

Rolling stock and crew scheduling are two critical and fundamental problems in the field of urban rail operations, especially under some flexible operation schemes and large-scale urban rail networks. These problems are often solved separately and sequentially from the perspectives of different planning stages, resulting in limited planning adjustment ranges of crew schedules and multiple feedback of two problems. To further make these two decision-making processes consistent, this paper investigates a joint rolling stock and crew scheduling problem in urban rail networks. A novel optimization model is formulated with the aim of reducing the operational cost of rolling stock units and crew members simultaneously, in which the multiple-train composition mode is also taken into consideration to adequately match dfferent frequency requirements and rolling stock transport capacities. To solve the proposed model, a customized branch-and-price-and-cut approach is designed to find the optimal operation schemes, in which Benders decomposition is used to solve the linear programming relaxation of the path-based reformulation. Two customized column generation methods with label correcting are embedded to solve the master problem and sub-problem for generating promising paths (columns) associated with each rolling stock unit and crew group, respectively. Finally, a branch-and-bound procedure is used to find integral solutions with several acceleration techniques. To demonstrate the performance and the robustness of the proposed approaches, a series of numerical experiments are conducted both on small-scale and real-life cases of the Beijing MTR urban rail network with different variations in the problem instance settings.

4 - Integrated Optimization of Train Makeup Problem and Resource Scheduling in Railway Shunting Yards: A Hybrid MILP-CP Approach with Logic-Based Benders Decomposition

Peiran Han, Beijing Jiaotong University, Beijing, China, People's Republic of, Lingyun Meng, Nikola Besinovic, Xiaojie Luan

In the shunting yard, various complex operations occur, leading to inefficiencies in railcar connections. Therefore, designing an effective operational research methodology is essential for the shunting yard, and even for the local rail freight network. At the planning level, this paper addresses the integrated train makeup problem and resource scheduling problem (TMRS). The train makeup is treated as an assignment problem, guiding all operations while a series of hybrid flow shop scheduling tasks are established to coordinate the operations of trains, blocks, and railcars. Due to the complexity of TMRS, the integrated problem is reformulated as a hybrid mixed-integer linear programming (MILP) and constraint programming (CP) model. Logic-based benders decomposition (LBBD) is used to partition the TMRS problem, with lower bounds designed and integrated into the solving procedure to accelerate the convergence. We propose feasibility cuts, optimality cuts, and symmetry cuts based on the structure of the subproblem, which are dynamically added to the master problem. Two numerical examples are designed to demonstrate the effectiveness of the proposed hybrid method and symmetry cuts. Finally, the proposed approach and algorithm are examined on a series of artificial instances and real-scale examples, demonstrating their practical effectiveness and ability to achieve high-quality solutions.

SB40

Summit - 432

New Topics in BOM

Invited Session

Behavioral Operations Management

Chair: Ruth Beer, Baruch College, CUNY, New York, NY, United States

1 - Hasty Expectations: Judgment Bias in Process Analysis from Misappreciation of Jensen's Inequality

Jordan Tong, University of Wisconsin Madison, Madison, WI, United States, Daniel Feiler

How many weeks, on average, will it take your company to complete 30 client cases? When there is variability in a process capacity metric (e.g., some cases are easier than others), we argue that people tend to first average over that variability before performing other calculations required to answer such a question (e.g., division). Such "hasty expectations" sometimes works but often leads to predictable bias due to Jensen's inequality. In particular, because 1/x is a convex function, hasty expectations biases estimates when the units of the observed data and the estimate are inverses, i.e., one is a capacity rate (units per time) and the other is a cycle time (time per unit). Across five pre-registered experiments (including samples of managers and MBA students), we test the extent to which this mechanism drives predictable bias in managerial judgment tasks commonly assigned in core MBA classes, including bottleneck analysis, breakeven analysis, and project completion estimates. We show how the bias can be avoided by first converting the data’s units to be congruent with the requested managerial judgment, ensuring that hasty expectations does not introduce bias.

2 - Driving with Algorithms: The Role of Precision on Improving Human Decision-Making

Park Sinchaisri, University of California Berkeley, Berkeley, CA, United States, Philippe Blaettchen

Many everyday tasks require sequential, interlinked decisions that make it difficult for workers to identify the best strategy. As a result, Artificial Intelligence (AI)-based recommendation systems are applied increasingly to reduce complexity. However, due to human biases, there remains a gap between what is technically possible and what is utilized. This work seeks to understand how recommendation systems can help induce efficient behavior in sequential decision-making. We develop a sequential decision-making task in the form of a virtual electric vehicle driving game. Participants, equipped with algorithm-generated recommendations, need to make sequential charging decisions while facing uncertain traffic. Our experimental results offer key insights into how humans make sequential decisions and respond to different types of recommendations, both immediately and in terms of long-term learning. Based on these insights, we develop a recommendation system that takes into consideration possible deviations by drivers, with the objective of improving their long-term performance.

3 - A Behavioral Study of Self-Other Adoption Discrepancies in XAI

Zezhen He, Simon Business School, University of Rochester, Rochester, NY, United States, Fernanda Bravo, Yaron Shaposhnik, Leon Valdes

Despite major recent developments in machine learning (ML), there is ample empirical evidence that users do not always follow ML recommendations. However, the question of whether the subject that is affected by the prediction impacts adherence is not well understood. This question is relevant as many ML applications, from medicine to the judicial system to lending decisions, are made by a user but affect a third party. In this work, we conduct behavioral experiments to study whether the party that is affected by a user’s decision (the user, or another participant) impacts the adoption of ML recommendations. In addition, we explore whether the presence vs. absence of ML explanations moderates our results. We find no difference in adoption without explanations. In the presence of explanations, however, users are more willing to use ML when they are the affected party.

4 - Algorithm Reliance and Preferences for Information

Ruth Beer, Baruch College, CUNY, New York, NY, United States, Daniela Saban

We study individuals’ preferences for delegating managerial decisions to a matching algorithm, after a large set of alternatives has been screened and the best candidates have been identified. Two behavioral considerations are central to our setting: algorithm reliance (i.e., individuals' willingness to delegate decision rights) and preferences for information (i.e., individuals' preferences for making a more informed decision and learning).

SB41

Summit - 433

Queueing Theory and Applications

Invited Session

Applied Probability Society

Chair: Daniela Hurtado Lange, Northwestern University, Evanston, IL, United States

1 - Multi-Dimensional State Space Collapse in Non-Complete Resource Pooling Scenarios

Ellen Cardinaels, Eindhoven University of Technology, Eindhoven, Netherlands, Sem Borst, Johan S. H. van Leeuwaarden

The this talk we focus on an explicit multi-dimensional state space collapse (SSC) for parallel-processing systems with arbitrary compatibility constraints between servers and job types. This breaks major new ground beyond the SSC results and queue length asymptotics in the literature which are largely restricted to complete resource pooling (CRP) scenarios where the steady-state queue length vector concentrates around a line in heavy traffic. The multi-dimensional SSC that we establish reveals heavy-traffic behavior which is also far more tractable than the pre-limit queue length distribution, yet exhibits a fundamentally more intricate structure than in the one-dimensional case, providing useful insight into the system dynamics. In particular, we prove that the limiting queue length vector lives in a K-dimensional cone of which the set of spanning vectors is random in general, capturing the delicate interplay between the various job types and servers. For a broad class of systems we provide a further simplification which shows that the collection of random cones constitutes a fixed K-dimensional cone, resulting in a K-dimensional SSC. The dimension K represents the number of critically loaded subsystems, or equivalently, capacity bottlenecks in heavy-traffic, with K=1 corresponding to conventional CRP scenarios. Our approach leverages probability generating function (PGF) expressions for Markovian systems operating under redundancy policies.

2 - Finite-TIME Convergence to Stationarity of $M/M/N$ Queues: A Lyapunov-Poincar\'E Approach

Hoang Nguyen, Georgia Institute of Technology, Atlanta, GA, United States, Sushil Mahavir Varma, Siva Theja Maguluri

Service systems like data centers and ride-hailing are popularly modeled as queueing systems in the literature. Such systems are primarily studied in the steady state due to its analytical traceability. However, almost all applications in real life do not operate in a steady state, and so, there is a clear discrepancy in translating theoretical queueing results to practical applications. To this end, we provide a finite time convergence for a simple $M/M/n$ queue, providing a stepping stone toward understanding the transient behavior of more general queueing systems. $M/M/n$ queue exhibits a phase-transition at the so-called Halfin-Whitt regime. We obtain a tight characterization of the finite time queue length distribution in the super-Halfin-Whitt regime, all the way until the phase transition at Halfin-Whitt. We also obtain a finite-time bound for the distance to stationarity for the sub-Halfin-Whitt regime as well, along with other finite-time statistics such as mean queue length and tail bound.

To prove such a result, we employ the Lyapunov-Poincar\'e approach, where we first carefully design a Lyapunov function to obtain a negative drift outside a bounded set. Within the bounded set, we get a handle on the behavior using a local version of the canonical path method to obtain a local Poincar\'e inequality. A key aspect of our methodological contribution is in obtaining tight guarantees in these two regions, which when combined give us tight mixing time bounds. Moreover, our approach has the potential to be generalized to other settings beyond $M/M/n$ systems.

3 - The Effect of Removing the Four-Hour Access Standard in The Ed: A Retrospective Observational Study

Tomas Momesso, UCL School of Management, London, United Kingdom, Bilal Gokpinar, Rouba Ibrahim, Adrian Boyle

Time-based targets are used to improve patient flow and quality of care within Emergency Departments. While previous research often highlighted the benefits of these targets, some studies found negative consequences of their implementation. We found that lifting the four-hour standard was associated with a drop in short-stay admission and an increase in the average length of stay in the ED.

4 - Online Learning for Reward-Driven and Congestion-Aware Service Rate Control

Wenqian Xing, Stanford University, Palo Alto, CA, United States, Vahid Sarhangian, Yue Hu, Anand Kalvit

We study service systems with heterogeneous customer and server types, where different types of servers can cooperate to serve a single customer. This collaborative service mode generates higher rewards, but faces the tradeoff of potentially taking capacity away from customers waiting in the queue. For such systems with flexible collaboration structures, our goal is to find a scheduling and collaboration policy that effectively minimizes revenue loss and waiting costs. We analyze this problem by considering both the long-run average and transient cost criteria. Our results provide insights into optimal strategies for initiating collaboration among servers.

SB42

Summit - 434

Generative Diffusion Models: Theories and Applications

Invited Session

Applied Probability Society

Chair: Renyuan Xu, University of Southern California, Los Angeles, CA, United States

Co-Chair: Wenpin Tang, Columbia University, New York, NY, United States

1 - Robust Fine-Tuning of Diffusion Models as Stochastic Control

Yinbin Han, University of Southern California, Los Angeles, CA, United States, Meisam Razaviyayn, Renyuan Xu

Fine-tuning generative models can be computationally intensive, particularly when accounting for uncertainty in human evaluations. To address this challenge, we investigate the robustness of fine-tuning diffusion models in the presence of a corrupted reward model. We frame the fine-tuning of diffusion models as an entropy-regularized control problem with linear dynamics, incorporating an uncertainty set over the reward. By leveraging recent advancements in reinforcement learning for linear-quadratic regulators and the mathematical structure of diffusion models, we develop efficient (robust) policy gradient algorithms and provide a convergence analysis. Additionally, we conduct numerical experiments to validate our theoretical results.

2 - Score Approximation, Estimation and Distribution Recovery of Diffusion Models on Low-Dimensional Manifolds

Zixuan Zhang, Georgia Institute of Technology, Atlanta, GA, United States, Kaixuan Huang, Minshuo Chen, Mengdi Wang, Tuo Zhao

Diffusion models have achieved state-of-the-art performance in various generation tasks. However, their theoretical foundations are not well understood. We study score approximation, estimation, and distribution recovery of diffusion models, when data are supported on an unknown low-dimensional manifold. Our results show that with a properly chosen neural network architecture, the score function can be both accurately approximated and efficiently estimated. Furthermore, the generated distribution based on the estimated score function captures the data geometric structures and converges to the data distribution with minimax optimal rate, which depends on the intrinsic dimension of the manifold, indicating that diffusion models can circumvent the curse of data ambient dimensionality. More importantly, our analysis makes steps towards understanding score functions in diffusion models: (1) we study the non-asymptotic behavior of score functions when time goes to zero; (2) we show that the geometry of manifolds impacts the learning of score functions, that is, distributions supported on manifolds with larger reach are easier to be learned by diffusion models.

3 - Accelerating Convergence of Score-Based Diffusion Models

Yuxin Chen, University of Pennsylvania, Philadelphia, PA, United States, Gen Li, Yu Huang, Timofey Efimov, Yuting Wei, Yuejie Chi

Score-based diffusion models, while achieving remarkable empirical performance, often suffer from low sampling speed, due to extensive function evaluations needed during the sampling phase. Despite a flurry of recent activities towards speeding up diffusion generative modeling in practice, theoretical underpinnings for acceleration techniques remain severely limited. In this paper, we design novel training-free algorithms to accelerate popular deterministic (i.e., DDIM) and stochastic (i.e., DDPM) samplers. Our accelerated deterministic sampler converges at a rate O(1/T^2) with T the number of steps, improving upon the O(1/T) rate for the DDIM sampler; and our accelerated stochastic sampler converges at a rate O(1/T), outperforming the rate O(1/sqrt{T}) for the DDPM sampler. The design of our algorithms leverages insights from higher-order approximation, and shares similar intuitions as popular high-order ODE solvers like the DPM-Solver-2. Our theory accommodates ℓ2-accurate score estimates, and does not require log-concavity or smoothness on the target distribution.

4 - Convergence Analysis for General Probability Flow Odes of Diffusion Models in Wasserstein Distances

Xuefeng Gao, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong

Score-based generative modeling with probability flow ordinary differential equations (ODEs) has achieved remarkable success in a variety of applications. While various fast ODE-based samplers have been proposed in the literature and employed in practice, the theoretical understandings about convergence properties of the probability flow ODE are still quite limited. In this paper, we provide the first non-asymptotic convergence analysis for a general class of probability flow ODE samplers in 2-Wasserstein distance, assuming accurate score estimates. We then consider various examples and establish results on the iteration complexity of the corresponding ODE-based samplers. This is a joint work with Lingjiong Zhu.

SB43

Summit - 435

Design of Experiments: From A/B Testing to Adaptive Experimentation

Invited Session

Applied Probability Society

Chair: Chao Qin, Columbia University, Palo Alto, CA, United States

1 - Estimating Distributional Treatment Effects in Randomized Experiments

Shota Yasui, Cyberagent, Tokyo, Japan, Undral Byambadalai, Tatsushi Oka

When analyzing data from randomized experiments, one commonly used measure is the Average Treatment Effect (ATE). However, understanding the distributional treatment effects can often provide a richer perspective than solely focusing on overall average effects. Estimating the Distributional Treatment Effect (DTE) allows researchers to investigate how the treatment impacts different points in the distribution of the outcome variable, revealing the heterogeneity in treatment effects.

In this session, we introduce methods to estimate DTE, including a novel regression-adjusted approach to estimate a wide range of distributional parameters in the context of randomized controlled trials (RCTs). We estimate DTE in data from an RCT on an online platform. Our results deepen our understanding of user behavior in response to treatments.

2 - Adaptive Experimentation Methods at Amazon

Tanner Fiez, Amazon, Seattle, WA, United States, Lalit Jain, houssam nassif

This talk explores the challenges and solutions associated with applying adaptive experimental design (AED) methods in complex real-world scenarios. While AED techniques have shown promise in simpler, idealized settings, their behavior and guarantees are not well-understood in environments characterized by non-stationarity or with complex (e.g., heteroskedastic, endogenous) noise structures. We discuss such problems that arise in industrial settings, and present solutions based on sequential linear experimental designs.

3 - Dual-Directed Algorithm Design for Efficient Pure Exploration

Wei You, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Chao Qin

We consider pure-exploration problems in the context of stochastic sequential adaptive experiments with a finite set of alternative options. The goal of the decision-maker is to accurately answer a query question regarding the alternatives with high confidence with minimal measurement efforts. A typical query question is to identify the alternative with the best performance, leading to ranking and selection problems, or best-arm identification in the machine learning literature. We focus on the fixed-precision setting and derive a sufficient condition for optimality in terms of a notion of strong convergence to the optimal allocation of samples. By including the dual variables directly, we characterize the necessary and sufficient conditions for an allocation to be optimal. The use of dual variables allow us to bypass the combinatorial structure of the optimality conditions that relies solely on primal variables. Remarkably, these optimality conditions enable an extension of top-two algorithm design principle, initially proposed for best-arm identification. Furthermore, our optimality conditions give rise to a straightforward yet efficient selection rule, termed information-directed selection, which adaptively picks from a candidate set based on information gain of the candidates. We outline the broad contexts where our algorithmic approach can be implemented. We establish that, paired with information-directed selection, top-two Thompson sampling is (asymptotically) optimal for Gaussian best-arm identification, solving a glaring open problem in the pure exploration literature. Numerical experiments highlight the exceptional efficiency of our proposed algorithms relative to existing ones.

4 - Optimal Exploration is no harder than Thompson Sampling

Zhaoqi Li, University of Washington, Seattle, WA, United States, Kevin Jamieson, Lalit Jain

Given a set of arms and an unknown parameter vector, the pure exploration linear bandit problem aims to return the arm with the highest reward, with high probability through noisy measurements of the rewards. Existing (asymptotically) optimal methods require either a) potentially costly projections for each arm or b) explicitly maintaining a subset of the arms under consideration at each time. This complexity is at odds with the popular and simple Thompson Sampling algorithm for regret minimization, which just requires access to a posterior sampling and argmax oracle, and does not need to enumerate the set of arms at any point. Unfortunately, Thompson sampling is known to be sub-optimal for pure exploration. In this work, we pose a natural question: is there an algorithm that can explore optimally and only needs the same computational primitives as Thompson Sampling? We answer the question in the affirmative. We provide an algorithm that leverages only sampling and argmax oracles and achieves an exponential convergence rate, with the exponent being the optimal among all possible allocations asymptotically. In addition, we show that our algorithm can be easily implemented and performs as well empirically as existing asymptotically optimal methods.

5 - Closing the Computational-Statistical Gap in BAI for Combinatorial Semi-bandits

Po-An Wang, KTH Royal Institute of Technology, Stockholm, Sweden, Ruo-Chun Tzeng, Alexandre Proutiere, Chi-Jen Lu

We study the best arm identification problem in combinatorial semi-bandits in the fixed confidence setting. We present Perturbed Frank-Wolfe Sampling (P-FWS), an algorithm that (i) runs in polynomial time, (ii) achieves the instance-specific minimal sample complexity in the high confidence regime, and (iii) enjoys polynomial sample complexity guarantees in the moderate confidence regime. To the best of our knowledge, even for the vanilla bandit problems, no algorithm was able to achieve (ii) and (iii) simultaneously. With P-FWS, we close the computational-statistical gap in best arm identification in combinatorial semi-bandits. The design of P-FWS starts from the optimization problem that defines the information-theoretical and instance-specific sample complexity lower bound. P-FWS solves this problem in an online manner using, in each round, a single iteration of the Frank-Wolfe algorithm. Structural properties of the problem are leveraged to make the P-FWS successive updates computationally efficient. In turn, P-FWS only relies on a simple linear maximization oracle.

SB44

Summit - 436

Simulation Models for Intermodal Transportation and Future Mobility

Invited Session

Simulation Society

Chair: Xueping Li, University of Tennessee, Knoxville, Knoxville, TN, United States

1 - Empowering Simulation Modeling: An Automated Ontology Framework Enhanced by Large Language Models

Jose Tupayachi, UTK, Knoxville, TN, United States, Haowen Xu, Xueping Li

We present an automated ontology framework designed to generate simulation models, with support from large language models. This framework uses ontological techniques to represent concepts, relationships, and attributes. We showcase its effectiveness through a simulation case study.

2 - A Simulation-Enabled Optimization Framework for Intermodal Freight Transportation

Zeyu Liu, West Virginia University, Morgantown, WV, United States

With the rise of E-commerce, freight transportation has been playing an increasingly important role in the U.S. economy. However, unimodal transportation, which uses solely heavy-duty vehicles or locomotives, typically leads to higher costs, less flexibility, and excessive greenhouse gas (GHG) emissions. Although the benefit of multimodal transportation has been pointed out by many, it still remains challenging to jointly optimize service routes and container flows for intermodal transportation problems, especially in large-scale networks. In this study, we develop a novel simulation-infused optimization framework to address the computational complexity in intermodal transportation problems involving heavy-duty vehicles, locomotives, and container ships. Our method seeks to minimize the total operating cost associated with large volumes of freight, as well as the life-cycle GHG emissions from all modes of transportation. We design novel linear approximation techniques to hierarchically combine agent-based simulation with mixed-integer programming models, optimized by a decomposition-accelerated branch-and-cut algorithm. We collect data from nationwide intermodal transportation facilities to validate framework performances through comprehensive experiments. Sensitivity analyses are conducted under what-if scenarios to provide insights to practitioners.

3 - Drone Pickup and Delivery System for Smart Cities

Aliza Sharmin, University of Tennessee Knoxville, Knoxville, TN, United States, Xueping Li

In drone pickup and delivery systems, identifying the optimal service design is a major challenge. The existing models of implementing drone pickup and delivery services either consider another mode of transportation to be in parallel operation with drones or require a depot or distribution center as a fixed start or end point of drones. This study envisions an integrated framework that utilizes solely drones to pick up and deliver packages from any origin to any destination within a region enhancing autonomy and economics in logistics. Instead of a drone delivery network that operates for a specific facility resulting in more non-delivery flight time due to fixed start and end points, this research proposes an extensive drone network being utilized by multiple facilities which can be prevalent in a city. The proposed system takes several factors, such as drone range, drone fleet size, charging time, charging strategies into account determining the optimal service design. The simulation specifically evaluates the performance of different charging strategies. Additionally, the study analyzes the impact of charging time variation, demand scenarios, drone fleet size and range, providing a more comprehensive understanding of the system and the factors that affect its performance.

4 - Energy Transaction Optimization of Hydrogen-Fueled Charging Station with Local Building Prosumers

Yang Chen, Oak Ridge National Laboratory, Oak Ridge, TN, United States, Olufemi Omitaomu, Haowen Xu, Xueping Li

Advancements in two-way communication smart grid technologies and distributed renewable energy have spurred growing interest in energy trading among neighborhood prosumers. This trend has led to the development of local energy transaction markets. This study focuses on the dynamics of energy transactions within a charging station-centric local energy market. Here, the charging station operates as the leader, while community building prosumers act as followers. The charging station is equipped with solar energy and a comprehensive hydrogen storage system, including a hydrogen tank, electrolyzer, and fuel cell. To delve into the pricing strategy within this local energy market, we propose a leader-follower optimization model. The model's upper level aims to maximize the operation profit for the charging station, whereas the lower level focuses on minimizing bill payments for the building prosumers. These prosumers possess solar energy and battery storage systems and can adjust their demand in response to pricing signals. Different groups of experiments have been designed to explore the effects of diverse building loads, electric vehicle drivers' behaviors, and levels of uncertainty on local energy pricing and transaction behaviors. Our experimental findings indicate that greater cost savings and energy-sharing levels are achievable within more heterogeneous building clusters.

5 - Simulation-Based Decision Support for Intermodal Transport Optimization with Container Consolidation under Demand Uncertainty and Limited Carbon Emission

Xudong Wang, The University of Tennessee, Knoxville, Knoxville, TN, United States

As global concerns over warming intensify, logistics stakeholders face pressures to reduce greenhouse gas emissions and energy consumption. Intermodal transportation, involving multiple transfer points and transport modes, which is a strategic response to these environmental challenges. However, balancing logistics costs with emissions reduction becomes complex amid unpredictable shipment demands and the prerequisite to pre-schedule diverse transportation methods. This study addresses these challenges by developing a decision-support system for intermodal logistics, optimizing the use of containers and transportation modes. Our goal is to develop a decision-support system that not only accommodates the initial transportation schedule but also flexibly adapts to upcoming demands that could share routes or resources, to enhance container and transportation mode utilization. To achieve this, we apply simulation-based methods that model various scenarios of containers and transportation modes availability. This approach allows us to test and refine strategies that dynamically adjust routing and scheduling decisions based on new information, reducing redundant resource usage and optimizing operational efficiency. Our simulation framework considers factors such as transportation and container booking costs, delivery time constraints, carbon emission limitation, and the logistic company's capability to synchronize resources effectively. This solution provides logistics companies with a strategic tool that fosters more informed and flexible decision-making in an uncertain market. This tool aims to boost logistical efficiency and sustainability by maximizing resource use and minimizing environmental impact, benefiting logistics service providers focused on enhancing adaptability and strategic planning.

SB45

Summit - 437

Emerging Technologies in Healthcare

Invited Session

Health Applications Society

Chair: Vedat Verter, Smith School of Business, Queen's University, Kingston, ON, Canada

1 - Large Language Models for Patient Flow Predictions

Kimberly Villalobos Carballo, Massachusetts Institute of Technology, Cambridge, MA, United States, Liangyuan Na, Yu Ma, Leonard Boussioux, Cynthia Zeng, Luis Soenksen, Dimitris Bertsimas

Accurate predictions of patient outcomes can facilitate resource allocation and enhance personalized care. In collaboration with a large hospital network, we developed machine learning models that predict short-term and long-term outcomes for all inpatients using electronic medical records. We implemented an automated pipeline displaying our daily predictions with user-friendly software. Over 200 medical staff currently use our tool, resulting in a significant reduction in length of stay and projected annual benefits of millions of dollars for the healthcare system.

Given this successful implementation, the question arises: how could we extend these tools for the benefit of hospitals with limited resources, small patient populations, and/or non-standardized healthcare records? Even though electronic medical records are widely available for most digitized systems, tabular data in healthcare is generally disorganized, not standardized across institutions and scarce in small hospitals. We then identified two significant limitations in the existing approaches to handling tabular data: they require labor-intensive data processing, and they ignore contextual information which could be used for data augmentation.

To address these limitations, we present TabText, a systematic framework that leverages Large Language Models to extract contextual information from tabular structures, resulting in more complete and flexible data representations. These representations can then be used to train any standard machine learning model for downstream prediction tasks. We demonstrated the flexibility of our approach compared to traditional labor-consuming processing techniques, and we showed that TabText can significantly improve performance across all patient outcome classification tasks considered, especially those with small data sizes.

2 - Nurse Workload Balancing via Real-TIME Location Data

Vedat Verter, Queen's University, Kingston, ON, Canada, Pooya Hoseinpour, Beste Kucukyazici

Sustained inequities in the workload distribution can lead to increased stress and reduced job satisfaction, high turnover and shortages in the nursing team. We utilize an extensive data set collected by a real-time location system installed in the surgical services department of a large tertiary hospital. The nurse workload is modeled as a multi-attribute, multilinear function, where the significance of each attribute (for the nurse manager) is elicited using an inverse optimization procedure integrated into a clustering method. This involves inverse optimization with a nonlinear integer original problem, which has not been well studied in the literature. We also robustify the model to incorpo- rate the uncertainties in the attribute weights. Through a real-life case-study we show that the proposed approach would have reduced the maximum travel distance by 50% - 77%. In addition, the average ranges for total direct care, maximum travel distance, and number of assigned patients decrease by 65%, 45%, and 31%, respectively.

3 - Patient Scheduling Under No-Shows: An Application to Diagnostic Services

Elaheh Fata, Queen's University, Kingston, ON, Canada, Mina Dalirrooyfard, Yuriy Nevmyvaka, Vedat Verter

One of the challenges that medical facilities face is the no-show behavior of already booked clients. To reduce the doctor's idle time cost, schedulers overbook clients, while keeping in mind the waiting cost exerted on clients. The goal of the medical facility is to minimize the expected sum of clients' waiting cost, the doctor's idle and overtime costs. We first study the static problem, where the scheduler knows the set of clients and their characteristics in advance. Next, we study the online setting and propose algorithmic solutions for scheduling clients as they come in over time.

4 - Quantifying the Cost-Effectiveness of Whole Genome Sequencing Strategies

Sripad Devalkar, Indian School of Business, Hyderabad, India, Sarang Deo, Mayank Jha, Pankaj Jindal, Abhishek Reddy

We develop an integrated operational and disease transmission model that captures the operational characteristics of a whole genomic sequencing (WGS) based surveillance program to detect novel variants of concern of a pathogen. The operational model estimates (i) the time to detect a novel variant circulating in the population and implement a public health measure in response and (ii) the cost of implementing a WGS-based surveillance program. The transmission model estimates the cost of public health intervention and its effectiveness in population-level health outcomes, such as the number of infections and deaths averted and life-years saved. Using this model, we assess the incremental cost-effectiveness for a combination of different operational configurations of the surveillance network (proportion of positively tested samples sequenced, the capacity of the network, centralized vs. decentralized structure) and epidemiological characteristics (severity, immune escape properties, and the time of emergence of the novel variant).

Across the simulated scenarios, we find that increasing sequencing capacity is cost-effective when a large proportion of positive samples are sent for genomic sequencing. Increasing capacity is not advantageous at a lower sampling proportion of positive cases. Centralized configurations do not significantly differ in turnaround time of variant detection (TAT) from decentralized configurations except when the novel variant emerges very late and a large proportion of positive samples are sequenced Optimal cost-effective operational configuration varies based on variant characteristics and its emergence time.

SB46

Summit - 438

Drone Use in Healthcare

Invited Session

Health Applications Society

Chair: Shakiba Enayati, University of Missouri - Saint Louis, St Louis, MO, United States

1 - Shortening Emergency Medical Response TIME with Unmanned Aerial Vehicle-Ambulance Joint Operations

Nan Kong, Purdue University, West Lafayette, IN, United States, Xiaoquan Gao, Paul Griffin

Unmanned aerial vehicles (UAVs) can improve emergency medical service (EMS) logistics by quickly delivering medical interventions with the help of bystanders. We formulate a large-scale MDP model to jointly optimize the dispatching and redeployment of UAVs and ambulances in real time. To tackle the curse of dimensionality, we adopt an approximate dynamic programming approach with neural network-based value function approximations. We construct a set of basis functions based on queueing and geographic properties of the UAV-augmented EMS system to design the approximation. Our approach outperforms existing benchmarks and provides guidance for effectively incorporating UAVs into EMS operations. By highlighting the potential advantages of using UAVs, we hope to encourage their wider adoption in the EMS field.

2 - Integrated Planning and Control of Drone Networks for Emergency Medical Response

Jamal Chu, University of Toronto, Toronto, ON, Canada, Sheng Liu, Wei Qi, Timothy Chan

Drones have been proposed as a supplemental response to medical emergencies such as cardiac arrest and anaphylactic shock through the delivery of time-critical supplies. Recent work has focused on separate drone base placement and heuristic dispatch policies via two-stage models, which may lead to performance loss compared to an integrated model. We first propose a new dispatch policy and prove that it is optimal under the assumption that drones have non-overlapping coverage zones. We then propose a novel stochastic integer program to jointly optimize base placements and dispatch policies and compare our integrated placement-dispatch model to a two-stage model. We efficiently solve large scale instances of this model using L-Shaped decomposition and a separation subroutine based on linear Irreducible Inconsistent Subsystems.

3 - Optimization of Drone and Truck Operations for Disaster Relief

Ricardo Modrego, Georgia Institute of Technology, Atlanta, GA, United States, Sofia Perez-Guzman

In post-disaster scenarios, the distribution of relief supplies is a very complex process due to lacking the relief supplies and transportation means to completely and promptly satisfy survivors' needs. The recent technological development of unmanned aerial technologies positions drones as potentially advantageous technology to support disaster relief distribution processes. Despite their limited capacity to carry goods, drones can access remote areas and attain faster speeds than road vehicles typically used to deliver relief supplies. The study develops a mixed-integer non-linear model comprising the routing, location, and allocation decisions for the distribution of relief supplies to survivors. The mathematical formulation minimizes a social cost function comprising logistic and suffering costs. The former costs are associated with trucks and drones transporting relief supplies and setting up PODs. The latter costs are associated with survivors' mobility and deprivation. Supplies can be transported via multiple deliveries from distribution centers to points of distribution (PODs). Survivors have heterogeneous needs based on their vulnerability. Individuals who cannot walk to their preferred POD stay at their initial location and have the supplies delivered. Trucks can also carry drones to PODs, and drones can independently fly to the staying survivors. The formulation integrates multiple complex parts, such as the hybrid vehicle fleet, deprivation costs, and survivor mobility. This research contributes to disaster relief distribution operations by introducing a drone fleet to complement the traditional truck network and modeling mobile survivors' mobility trajectories, further prioritizing vulnerable survivors and reducing social costs.

4 - Resilient Delivery Logistics: Integrating Truck and Drone Planning with Disruption Response

Mehmet Kurt, Auburn University, Auburn, AL, United States, Alice Smith, Daniel Silva Izquierdo

This study investigates the potential impacts of disruption, including adverse weather and unforeseen events like component failures or electromagnetic interference, on the well-known Flying Sidekick Traveling Salesman Problem (FSTSP). It introduces the first FSTSP extension with a recourse for disruptions, examining total disruption, affecting the entire planning area for a fixed period, and rolling disruption, progressing sequentially.

A two-stage model is introduced. The first stage plans routes for both the drone and the truck at the start of the day, while the second stage devises routes for remaining nodes during drone disruptions. This is a reactive strategy that reroutes as needed in response to a disruption. Solutions are obtained through Mixed Integer Linear Programming (MILP-exact) and heuristic methods for small instances, and solely heuristic for larger ones, adapted from Murray & Chu (2015).

Small-scale tests revealed a 3.1% gap between the exact and heuristic methods, indicating heuristic efficacy. On average, truck-and-drone scenarios performed 11.5% and 9.1% better than truck-only cases for the 12-customer exact and heuristic methods, respectively. In the 100-customer scenarios, the improvement was 5.7%. The recourse action for the rolling disruption proved 6.2% more effective on average than for total disruption, while the effect of the total disruption (with recourse action) was 16.7% relative to no disruption. Furthermore, experiments with perfect information demonstrated significant enhancements over reactive planning, all as expected.

This study offers key insights for truck and drone delivery planning, emphasizing the importance of considering disruptions during drone usage for efficient last-mile operations.

SB47

Summit - 439

Data-driven Research on Healthcare Operations

Invited Session

Health Applications Society

Chair: Nan Liu, Boston College, Chestnut Hill, MA, United States

Co-Chair: Miao Bai, University of Connecticut, Storrs, CT, United States

1 - Cost-Effective Surveillance Strategies for Early Detection of Cancer Recurrence: Optimal Integration of Imaging and ctDNA Testing

Reza Skandari, Imperial College London, London, United Kingdom, Narges Mohammadi, Niloofar Zamani, Alexander Pearson

Recurrent cancer poses significant challenges, with fatalities often occurring within a year of detection. While standard practice relies on post-treatment radiological surveillance, emerging evidence suggests ctDNA, a novel blood test, may offer earlier detection. We aim to develop cost-effective strategies to identify the optimal choice and timing of testing modalities (imaging and ctDNA), minimizing recurrence detection time to improve survival while controlling screening costs. We utilize a partially observable Markov decision process to formulate optimal post-surveillance strategies. Recognizing the importance of the cost-effectiveness threshold, we aim to provide policymakers with a menu of optimal policies and their performances. We apply our developed modeling framework to head and neck cancer and utilize a data-driven model to assess alternative policies. We finally evaluate policy guidelines and recommendations against the optimal policy.

2 - Cost-effectiveness of a network of lethal ovitraps for the prevention and control and dengue fever

Yvonne Huijun Zhu, National University of Singapore, Singapore, Singapore, Joel Aik, Shuzhen Sim, Joel Goh

An estimated 390 million people around the world are infected with dengue fever each year, and this number is trending upwards. One of the interventions that the World Health Organization has recommended to control the spread of dengue fever is the use of a network of lethal ovitraps to capture dengue vectors: female Aedes mosquitoes that carry the dengue virus. Existing evidence supporting the use of this intervention has largely focused on the efficacy of these traps, but has ignored the substantial operational costs entailed in servicing the network of traps.

We introduces methodology to perform a system-level cost-effectiveness analysis of such an intervention. A novel feature of the method is that it incorporates elements of the operational processes required to maintain the ovitrap network and accounts for the labor costs in doing so. Effectiveness was measured as disability-adjusted life years (DALYs) averted from the reduced mosquito population due to the network of lethal ovitraps, modeled using a vector-host dynamic compartmental model. Labor costs were estimated by modeling the lethal ovitraps as nodes of an undirected graph, constructing approximately-balanced routes for service personnel that accounted both for the travel time between locations and the time taken to maintain each trap, and accounting for the costs of employing sufficient numbers of service workers to maintain all traps at a given service interval.

3 - Helping the Captive Audience: Advance Notice of Diagnostic Service for Hospital Inpatients

Zheng Zhang, Zhejiang University, Hangzhou, China, People's Republic of, Miao Bai, Nan Liu

Inpatients are often treated as the “captive audience” on-demand for hospital diagnostic service, and they are notified only when service capacity is available. This management approach, commonly employed by hospitals in both China and the United States, can cause significant chaos and inefficiencies. We propose an innovative scheduling approach called “advance notice” to manage hospital diagnostic practice. Advance notice is a brand-new scheduling paradigm in between the classic allocation scheduling and advance scheduling. Patients are placed in a common queue waiting to be called for service, and they will be provided both a fixed preparation time and a guaranteed service time window in advance (neither a last-minute notice nor an exact service time in the future). The advance notice policy enjoys the benefit of allocation scheduling and that of advance scheduling. It calls for two decisions: who to serve now and who to send advance notices to. We formulate a Markov Decision Process model to optimize these decisions dynamically. Via a novel variable transformation, we reveal the hidden antimultimodular structure of the problem and show how optimal decisions should be adjusted in response to changes in the system load. Beyond solving the MDP model for daily operations, we further investigate how the service time window and preparation time window, as system-level controls, can be used to manage system performance. Our numerical study demonstrates significant improvement in operational efficiency by switching from the current practice to adopting our proposed advance notice policy.

4 - Are No-Shows All That Bad? Strategic Double-Booking and When to Intervene

Arseniy Gorbushin, Rotman School of Management, Toronto, ON, Canada, Opher Baron, Olga Bountali, Binyamin Oz

No-shows constitute a long-standing challenge in various service sectors, particularly in healthcare, where they jeopardize resource utilization and patient care quality. Traditional approaches to reduce no-show rates, such as reminders, cancellation fees, and flexible scheduling have limited effectiveness, overlooking the strategic nature of double-booking—a phenomenon where individuals deliberately secure multiple appointments to expedite service.

We model strategic double-booking, investigate its impact on system dynamics, and explore strategies to tame it. In particular, we use a stylized queueing model and a game-theoretic framework to incorporate the notion of strategic double-booking. The system that we study consists of identical queues (e.g., hospitals), where customers have the option to double-book (or multi-book) upon arrival. We show that the model exhibits interdependencies among queues, that are intrinsic in the system dynamics. We analytically account for both them and the resulting capacity waste, providing insights into customer decision-making and system performance.

Our analysis reveals that double-booking has dual effects: Facilitating load distribution while increasing system traffic.
We show that increasing the perceived value of service (i.e., reward) does not always lead to a higher number of served customers, which is the typical result in the strategic queue literature. In addition, we prove that banning double-booking may not always improve system performance, especially under low traffic conditions. We also demonstrate how the presence of double-booking impact traditional remedies, such as no-show penalties and overbooking. Finally, we propose alternative solutions, such as a flexible admission control policy, to allow load distribution without excessively congesting the system.

SB48

Summit - 440

Revenue Management in Supply Chain

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Bin Hu, University of Texas at Dallas, Richardson, TX, United States

Co-Chair: Haokun Du, The University of Texas at Dallas, Richardson, TX

1 - Competitive Markovian Pricing

Haokun Du, The University of Texas at Dallas, Richardson, TX, United States, Bin Hu, Elena Katok

Problem definition: Dynamic pricing is a popular price-discrimination strategy in retail; however, it is often undermined by customers' strategic waiting behavior. One tactic utilized by retailers to deter strategic waiting is to offer price discounts at random intervals. This phenomenon has been studied in the recent Markovian-pricing literature in single-retailer settings. In this paper, we extend the study to a competitive setting where two retailers offer random price discounts for incoming customers, and also investigate the impact of customer loyalty in this setting. Methodology/results: We perform an equilibrium analysis of a competitive Markovian-pricing model with loyal and opportunistic customers and recover a key insight from the single-retailer Markovian-pricing literature that the retailers should offer flash discounts that last only for an instant. Surprisingly, we also find that a larger proportion of opportunistic customers at both or each of the retailers may increase both retailers' profits when they engage in competitive Markovian pricing. Managerial implications: Our study extends the support for the use of flash discounts to competitive retail settings. The surprising finding that more opportunistic customers in the market or at each retailer may benefit both competing retailers highlights the unique properties of the competitive Markovian-pricing setting where common wisdom may not apply, and suggests that retailers engaging in competitive Markovian pricing should reconsider their policies regarding price-monitoring and comparison services as well as their efforts to foster customer loyalty.

2 - Familiarity-Based Dynamic Pricing with Hierarchical Bayes Estimation of Consumer Choice

Binghan Kou, Arizona State University, Tempe, AZ, United States, Hongmin Li, Yuqi Yang

Past consumption affects customers' familiarity with a product and influences their preference for future consumption of the product. This effect is heterogeneous among individuals. More generally, individuals' preferences may be nonmonotonic with respect to familiarity and exhibit various patterns. Meanwhile, the prevalence of customer management programs has made consumer transactional data readily accessible. This creates opportunities for the firm to tailor pricing decisions to each customer's specific familiarity-based utility pattern. A major hurdle, however, is the complexity of deciphering the diverse familiarity-based utility patterns embedded in the transactional data and integrating this information with the price optimization model. In this paper, we present a value-passing dynamic pricing strategy that is optimal under a familiarity-based multinomial logit (MNL) choice model. In this context, we develop several flexible variants of the familiarity-based pricing model that dramatically reduce the complexity of integrating diverse utility patterns into optimal pricing decisions. Employing a hierarchical Bayes estimation approach, we demonstrate how this data-driven dynamic pricing strategy improves a firm's bottom line.

3 - On the Value of Coalition Loyalty Programs

Jingmai Wang, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, Yan Liu, Yulan Wang, Dan Zhang

A growing number of firms join forces with other firms in their loyalty program effort by participating in coalition loyalty programs (CLPs). Despite their increasing popularity and importance, CLPs remain significantly understudied, especially compared to proprietary loyalty programs (PLPs) offered by individual firms. This paper studies CLPs and compares them to PLPs using an analytical framework. We model a CLP consisting of n firms, each of which sells a nondurable product to infinitesimal, heterogeneous customers over an infinite time horizon. Our research reveals the critical role of market composition and customer discounting on the effectiveness of CLPs. We show that larger-sized CLPs are only preferred in markets where customer heterogeneity in valuation is more prominent than their heterogeneity in shopping intensity. Furthermore, when customers discount the future, smaller-sized CLPs can be more effective. Importantly, we show that firms have incentives to join CLPs even when PLPs are ineffective, corroborating the increasing popularity of CLPs in recent years. Our results offer useful guidelines for the design and operation of CLPs to firms and coalition managers. For example, we offer one explanation for the struggles of some CLPs in their expansion. We check the robustness of our results with model extensions and discuss the limitations of our study.

4 - Yielding to the Weather: How Climate Change Impacts Planting Decisions and Consumers

Yaniv Ravid, University of Toronto, Toronto, ON, Canada, Hongqiao Chen, Ming Hu, Wenbin Wang

Climate change is affecting crop yields worldwide and presenting threats to the global food supply chain. In this paper, we characterize the impact of climate change on crops' yields and analyze the downstream ramifications these changes have on consumers and agricultural firms' planting decisions. We assume that firms can influence prices by deciding on planting quantities in or across several growing regions, which will result in random crop yields depending on environmental conditions. As weather and climate trends vary across growing seasons, we analyze how changes to the yield distribution affect consumer surplus and the greater agricultural supply chain. We analyze both the cases of a monopolistic firm and competing firms and show that when the correlation in yield between agricultural regions decreases, consumer surplus under a monopoly always increases but may decrease in the case of competition. In addition, we show that competing firms may benefit from an increase in yield volatility, while a monopolistic firm's profit will always decrease after such changes to the yield distribution. We also show that an increase in average yield hurts competing firms in some instances, while a monopolistic firm always benefits from an increased average yield. These results show that a lack of coordination between competing firms could therefore be hurting consumers and firms alike. We illustrate our results and their implications on how firms should alter their planting decisions using practical examples of crops and regions. Our results provide guidance for agricultural operations on how to adapt to climate change.

SB49

Summit - 441

Operations and Society

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

1 - Human-Artificial Intelligence Teaming and the Effects of Workload on Child Welfare Screening

Alan Scheller-Wolf, Tepper School of Business, Pittsburgh, PA, United States, Yanhan Tang, Zhaohui Jiang, Justine Galbraith, Lindsey Lacey

Child welfare organizations regularly receive a significant number of calls alleging child neglect or abuse. Due to limited resources available for investigations and services, it is crucial to accurately assess and screen these allegations before further investigation or intervention to ensure service quality and efficiency. Furthermore, investigations initiated based on unsubstantiated allegations can lead to harmful consequences for the family involved. To aid these essential screening decisions and enhance overall efficiency, a Predictive Risk Model (PRM), essentially an artificial intelligence (AI) tool, has been deployed by our research partner. We empirically investigate this human-AI collaboration in call screening, particularly examining the influence of workload on this collaboration. We find that human agents are more likely to deviate from the AI recommendations when the workload is either high or low. While the AI tool does not adjust for varying workloads, human agents seem to factor in the workload when making screening decisions. Specifically, they lean toward admitting more low-risk referrals when the system load is low and lean toward rejecting more high-risk cases when the load is high, resulting in a U-shape relationship between deviation and workload. These findings indicate that human workers can complement the AI's recommendations by taking into account vital operational factors, such as system load. We provide evidence supporting that a sustainable investigation load is likely desirable for downstream operations, as increased investigation load is correlated with delayed investigation completion.

2 - Incentive Design for Sustainable Practices In Smallholder Supply Chains

Yuan Shi, Massachusetts Institute of Technology, Cambridge, MA, United States, Joann de Zegher, Yanchong Zheng

Promoting sustainable practices by smallholder farmers face significant challenges of resource constraints and uncertain crop prices. In practice, sustainability incentives are provided via two main mechanisms: (i) an offsetting scheme in which third-party nonprofits or project developers provide financial rewards for adoption; (ii) an insetting scheme in which downstream buyers directly incentivize farmers within their value chain. It is an open debate as to when either scheme may be more effective and economical than the other to promote long-term adoption of sustainable practices among smallholder farmers. In this paper, we examine optimal incentive design and economic viability for these two schemes when farmers are resource constrained and face price variability for their products. We develop a principal-agent model in which the principal aims to ensure continued compliance with sustainable practices at minimal cost, while farmers repeatedly make joint decisions on production effort and sustainability compliance over multiple periods. Our analysis suggests that insetting contracts with incentive payments depending on both compliance and production volume can significantly outperform offsetting contracts which focus exclusively on sustainability compliance. This is because insetting contracts help to prevent farmers from engaging in undesirable tradeoffs between production intensification and sustainable actions. We further examine and demonstrate the near-optimal performance of affine insetting contracts which are more implementable in practice. Using field data of oil palm farmers in Indonesia, we quantify that well-designed insetting contracts can achieve over 35% cost savings than offsetting contracts, while achieving the same level of sustainability.

3 - Child Welfare Platform Design to Improve Outcomes for Children with Disabilities

Ludwig Dierks, University of Illinois at Chicago, Chicago, IL, United States, Vincent Slaugh, Utku Unver

Children with disabilities age out of the U.S. child welfare system without a family resource at disproportionately high rates, resulting in worse life outcomes. Child welfare platforms seek to serve these children but struggle to identify high-capability families and direct them to children with high needs. Moreover, families without the necessary skills and resources for children with disabilities might express an eagerness to adopt any child, failing to recognize their own limitations. This costs agencies valuable screening resources and undermines the platform's usefulness for caseworkers.

We model the foster care or adoption placement process for a platform serving low-needs and high-needs children and low-capability and high-capability families. The platform elicits each family's type, which is unobservable. We derive closed-form expressions to characterize three placement mechanisms: a naive maxima mechanism that aims to maximize the number of placements but neglects families' incentives, a maximal envy-free mechanism that maximizes the number of placements while accounting for families' incentives, and a max-min envy-free mechanism that puts extra importance on helping high-needs children. The superior performance of the envy-free mechanisms demonstrates the importance of accounting for incentives.

Finally, we discuss managerial insights. We show that the ability to place high-needs children with high-capability families depends on the ratio of low-needs children to families unable to care for high-needs children. This ratio affects three strategic decisions: recruiting more families of different care capabilities, including more low-needs children on the platform, and coaching different types of families to better understand their care capability.

SB50

Summit - 442

Responsible Operations and Workforce Management

Invited Session

MSOM: Healthcare

Chair: Hummy Song, Wharton, Philadelphia, PA, United States

Co-Chair: Harriet Jeon, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

1 - Redesigning Volunteermatch's Recommendation Algorithm: Toward More Equitable Access to Volunteers

Scott Rodilitz, UCLA Anderson School of Management, Los Angeles, CA, United States, Vahideh Manshadi, Daniela Saban, Akshaya Suresh

In collaboration with VolunteerMatch (VM)---the world's largest online platform for connecting volunteers with volunteering opportunities---we designed and implemented a new display ranking algorithm. VM's original ranking algorithm intended to maximize efficiency (i.e., the total number of connections), but as a consequence it repeatedly displayed the same few opportunities atop its ranking, effectively limiting access to volunteers for the other opportunities. To incorporate VM's desire for equity (defined as the weekly number of opportunities with at least one connection) along with efficiency, we propose a modeling framework for online display ranking in settings where it is important to manage the trade-off between the total number of connections and the equitable allocation of these connections. Taking an adversarial approach in evaluating the performance of online algorithms, we show that a class of algorithms that applies a *penalty* to opportunities after each connection provides a strong (and, in certain regimes, optimal) performance guarantee. Based on our theoretical development, we propose *SmartSort*, a simple online display ranking algorithm with a penalty term that we calibrated using VM's data and simulation. We implemented SmartSort in experiments covering Dallas-Fort Worth and all of Southern California. Using a difference-in-differences analysis, we find that the implementation of SmartSort led to an estimated 8% increase in equity (consistent across both experiments) without any significant decrease in efficiency.

2 - Stopping the Revolving Door: An Empirical and Textual Study of Crowdfunding and Teacher Turnover

Samantha Keppler, University of Michigan, Ann Arbor, MI, United States, Jun Li, Andrew Wu

Though the effectiveness and efficiency of education systems depends in part on a committed teacher workforce, high rates of teacher leaving have plagued the US public schools. A primary reason is that teacher workloads are high and involve many non-teaching tasks, such as procuring much of their own supplies and materials. This has recently become more salient with the emergence of teacher crowdfunding platforms. With the understanding that securing resources through crowdfunding may impact teacher work(loads), this paper investigates whether and how teacher crowdfunding affects teacher retention. We find that teachers funded on the largest teacher crowdfunding platform, DonorsChoose, are 1.6 percentage points (pp) less likely to leave their schools and 1.9 pp less likely to leave the teaching profession---a 14% and 41% reduction, respectively. The effect is larger for classroom environment (i.e., furniture) projects (4.2 and 2.8 pp) and for more unique projects (2.1 and 1.9 pp), evidencing that teachers are most impacted when they leverage the platform for resources particularly hard to get otherwise. This suggest that the effects are driven not only by the secured resources, but also by the efficient and autonomous platform procurement process. Improved US educational supply operations are needed to remedy the persistent gap between what resources teachers want and what they have. While that gap persists, teacher crowdfunding platforms appear to be a critical tool for stemming teacher turnover and attrition.

3 - To Each Their Own (Shifts): Incorporating Heterogeneous Worker Preferences into Shift Work Schedules

Hummy Song, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Harriet Jeon, Song-Hee Kim, Kyeongsug Kim, Sangwoon Cho, JeongHee Hong

While shifts are the dominant and often unavoidable way to schedule work in many contexts requiring around-the-clock staffing, their adverse effects, both at the individual and organization levels, are well-documented. We explore a potential organizational lever—incorporating heterogeneous preferences over shift design—to mitigate ramifications of shift work in the context of acute care bedside nurses. Using a novel combination of survey, administrative, and shift data, we examine whether, to what extent, and for whom individual choice over dimensions of their shifts mitigates the impact of shift work on work (dis)satisfaction and turnover of nurses.

4 - Policing Staffing and Fairness: Evidence from Atlanta Police

Jonathan Zhou, Georgia Institute of Technology, Atlanta, GA, United States, Qiuping Yu, Zheng Dong, Yao Xie

We study how scheduling and dispatching policies impact policing effectiveness and fairness using event-level police patrol and 911-call data collected across a five-year period from the Atlanta Police Department. We first empirically identify two major factors --- police's scheduling policy and the geographic division of the police patrolling units --- that lead to significant disparities in police response time to 911 calls across different neighborhoods. We then propose strategies to mitigate disparities while also improving overall policing outcomes using a framework that combines optimization and causal inference.

SB51

Summit - 443

Empirical Studies in OM and Finance Interface

Invited Session

MSOM: iForm

Chair: Qi Wu, Case Western Reserve University, Cleveland, OH, United States

Co-Chair: Sridhar Seshadri, University of Illinois, Champaign, IL, United States

1 - Effects of Financial Constraints on Supply Chain Financing Choices and Operational Decisions

Anqi Wu, Florida International University, Miami, FL, United States, Qi Wu, Sridhar Seshadri

Bank credit access represents the opportunity to borrow funds from banks and other financial institutions. Although theories have been developed on how bank financing affects firms' operational decisions, limited operations management literature empirically investigates this impact and the mechanisms through which it manifests. This study focuses on firms' inventory decision and employs the adoption of interstate bank branching laws that introduced a staggered shock on operational firms' accessibility to bank credit. We collate a panel dataset for 1990--2005 by merging several datasets with credit line information obtained by data mining 10-K filings. We examine the effect of bank credit on inventory decisions using a difference-in-differences identification strategy. Our study reveals that enhanced access to bank credit results in quicker inventory turnover rather than a rise in firms' inventory investments. This effect is more pronounced among small firms. We observe that in the short term, small firms focused more on capacity expansion, which may crowd out inventory investment. In the long term, their fast inventory turns can be attributed to improved infrastructure and higher capital intensity. For bigger firms, the benefits of relaxed bank credit constraints become evident through their influence on the supply chain network. We observe that when customer firms have more bank credit, larger suppliers are willing to provide additional trade credit, and this is associated with higher trade with the new customer and therefore faster inventory turns. This research contributes to a better understanding of the comprehensive effect of bank credit on firms' management of their operations.

2 - Impact of Supply Chain Transparency: Evidence from Conflict Minerals Disclosures

Christopher Chen, Indiana University Kelley School of Business, Bloomington, IN, United States, Paola Martin

Supply chain transparency has become crucial due to societal expectations, environmental concerns, and regulatory requirements, emphasizing both ethical sourcing and operational challenges. This study investigates how mandated transparency regarding conflict minerals impacts buyer-supplier relationships, specifically examining supplier base concentration and supplier financing. Using a game theoretic model, we capture strategic interactions between buyers and suppliers, highlighting that suppliers may consolidate to be labeled as conformant and to drive out competition. To empirically analyze transparency's impact, we leverage the Dodd-Frank Act's Section 1502, which mandates firms to disclose the origins of their raw materials, focusing on tin, tungsten, tantalum, and gold (3TG) from conflict regions. We compare firms subject to these requirements with those that are not, using hand-collected conflict mineral disclosures from SEC filings and Compustat data (2007-2021). Our generalized difference-in-differences model reveals that mandated transparency leads to a significant increase in trade credit granted to disclosing firms, suggesting improved financing terms. Additionally, there is a trend of supplier base consolidation among disclosing firms, impacting non-conformant suppliers. This consolidation results in more buyers per supplier, indicating a shift towards fewer, more compliant suppliers. These findings suggest that transparency fosters compliance and ethical sourcing while influencing supply chain strategy and financial relationships. Firms can leverage transparency to enhance their brand and gain competitive advantages through better supplier terms and efficient supply chain management. However, achieving transparency requires balancing the benefits of disclosure with operational and financial burdens, and it may reshape market dynamics, potentially squeezing out smaller suppliers.

3 - Stock Market Reaction to New Technology Announcements: An Empirical Investigation into the Role of Continencies

Ujjal Mukherjee, University of Illinois, Urbana-Champaign, Champaign, IL, United States, Alok Raj, Samit Paul, Atanu Chaudhuri

In this paper, we analyze the reaction of stock markets in the context of new technology adoption in supply chains. The specific technology context we use is blockchain technology. We demonstrate empirically that while stock markets generally react positively to new technology adoption announcements, several contingencies are important considerations. The signals that investors perceive from new technology adoption announcements depend significantly on the context and contingencies of announcements. Our paper is positioned at the intersection of supply chain management and finance.

4 - Customer Portfolio Approach to Managing Cash Flow Variability

Nikolay Osadchiy, Emory University, Atlanta, GA, United States, William Schmidt, Jing Wu

Cash flow variability is driven by operational decisions, and it influences operating performance and valuation. Despite this, and the early role that operations management scholarship had on cash flow management theory, the prevailing approaches for managing cash flow variability focus on financial remedies such as accessing external capital and engaging in financial hedging, and to a lesser extent, operational remedies such as sourcing goods in currencies that align with revenues and implementing product or production flexibility. Our research extends this set of remedies by proposing customer portfolio management and selective trade credit as operational hedges for reducing cash flow variability. We empirically validate our proposal by studying new customer acquisition events using a large database of customer-supplier relationships that we join with quarterly firm financial reports. Our analysis shows that firms reduce their cash variability by (i) pursuing customers with desirable order patterns that offset the cash flow variability from serving legacy customers and (ii) selectively offering customers trade credit that harmonizes payment terms in the customer portfolio. We strengthen the inferences from this analysis by assembling random samples of customers for each firm in our study, and show that the firms' actual customer portfolios yield lower cash variability compared to the counterfactual sets of randomly sampled customers.

5 - The Role of Firm Risk in Supply Chain Network Adaptation

M. Hakan Hekimoglu, Rensselaer Polytechnic Institute, Troy, NY, United States, Shailesh Divey, T Ravichandran

The paper addresses how firm risk affects its centrality within a supply chain network. Using supplier-customer relationship data collected from FactSet pertaining to the US manufacturing industry, we examine the effects of a firm's risk on its centrality. We measure a firm's centrality using two alternative constructs: Degree centrality and eigenvector centrality. We measure firm risk with the idiosyncratic volatility from the Fama-French three-factor model with industry effects. Our paper presents three important findings. First, we show that a firm's risk has a negative impact on its centrality in the supply network. A higher risk firm has fewer supplier/customer links (i.e., lower degree centrality) in addition to those suppliers/customers being less-central firms (i.e., lower eigenvector centrality). Second, we show how a firm's centrality adapts to a change in its risk exposure. Using a first-difference estimator, we find that an increase in firm risk leads to a decrease in the firm's eigenvector centrality, but it does not affect its degree centrality. We interpret this interesting supply chain network adaptation as follows: Customers and suppliers of firms experiencing higher risk cut their ties with the focal firm, so the focal firm has to replace those customers and suppliers with less central ones. Third, we present that a firm's centrality changes more dramatically in response to an increase in risk than it does to a decrease in risk.

SB52

Summit - 444

Technology and Future of Work

Invited Session

MSOM: Service Operations

Chair: Qiuping Yu, Georgetown University, Washington, DC

1 - The Effect of Income Composition on Consumption Behavior

Paige Tsai, Harvard Business School, Boston, MA, United States, Ryan Buell

In this work, we study whether the scheduling flexibility afforded by part-time working arrangements hinder the financial flexibility of employees by altering the composition of their income. To the extent that the composition of an employee's income wields influence over how they spend it, different working arrangements with the same total level of compensation may be differentially sustainable – with meaningful implications for employee wellbeing and operational performance.

Our preliminary evidence suggests that income composition influences consumption behavior, such that individuals whose income is more concentrated around their primary income source report spending more of their total income overall – an effect driven by purchases denoted as relative necessities. We believe this research will lay the foundation for interventions that could improve the design of operations to bolster employee financial wellbeing by understanding the role a job plays in the overall composition of an individual's income.

2 - On Worker Scheduling Fairness: Evidence from An American Restaurant Chain

Qiuping Yu, Georgetown University, Washington, DC, DC, United States, Xiaoxuan Hou

Utilizing real-world shift data from an American casual dining chain, we examine racial and gender disparities in scheduling quality conceptualized along three dimensions: sufficiency (whether workers are allocated enough hours as they desire), predictability (the frequency of short-notice and real-time changes to the schedule), and hourly earning potential (whether workers' scheduled hours translate into equitable hourly earnings). We investigate two main questions: (1) are there racial and gender disparities in the schedule quality among workers? (2) what are the mechanisms that drive the disparities and what are the mitigation strategies? Our findings indicate that compared to white male colleagues, female and minority workers are scheduled for hours with lower income potential and are more likely to receive real-time additions to their schedules. They also experience a higher proportion of hours added on short notice rather than pre-scheduled. Further, female workers are scheduled for fewer hours than male workers. By ruling out worker preferences and skill differences, our analysis points to management bias as one of the primary drivers of these inequalities.

3 - The Impact of AI Technology on The Productivity of Gig Economy Workers

Serguei Netessine, The Wharton School, Philadelphia, PA, United States, Dmitry Mitrofanov

We conducted field experiments on a grocery shopping platform that uses an AI-enabled guidance system to help shoppers find products on store shelves. We find that, as one would expect, such technology helps by reducing the number of refunds (due to the item being hard-to-find, e.g. located at a pop-up display, etc.) and that less experienced shoppers tend to use this guidance the most. However, counter-intuitively, we also find that the usage of more complex routing algorithms is not free: it takes a longer time to consult AI guidance and picking times increase as a result. Overall, we find that AI improves the effectiveness of gig workers by helping less experienced workers achieve order outcomes that are more comparable to those of more experienced workers, thus increasing both customer satisfaction and revenue per order. However, there are boundary conditions for technology adoption and overuse of technology can even lead to lower productivity.

4 - Measuring Strategic Behavior by Gig Economy Workers: Multihoming and Repositioning

Gad Allon, University of Pennsylvania, Philadelphia, PA, United States, Daniel Chen, Kenneth Moon

Gig economy workers make strategic decisions about where and when to work. We empirically measure two types of strategic behavior: multihoming, an online change between platforms, and repositioning, a physical change between locations. Using a structural model, we show that workers are highly heterogeneous in their preferences and find multihoming especially costly, both in absolute terms and relative to the cost of repositioning. Through counterfactual simulations, we show the significance of enabling free multihoming and efficient repositioning for improved worker earnings and platform service levels, providing policy insights for platform operators and regulators.

SB53

Summit - 445

Supply Transparency, Food Waste, and Carbon Emissions

Invited Session

MSOM: Supply Chain

Chair: Jing-Sheng Song, Duke University, Durham, NC, United States

Co-Chair: Chenghuai Li, Duke University, Durham, NC, United States

1 - An Empirical Study of Blockchain-Driven Transparency in A Consumer Marketplace

Wedad Elmaghraby, UMD, College Park, MD, United States, Ken Moon , Jane Jiang

Marketplaces have long been shaped by technological advancements from railroads to the mobile Internet and search. Now, with the potential to transform marketplace transparency, blockchain technology lets firms credibly trace the flow of products through their supply chains. It has been claimed that blockchain will soon enable consumers to track their food consumption from farm to table, uncover counterfeits and contamination, and better identify socially responsible products and firms. However, implementing a blockchain-based solution for supply chain transparency can be costly and complex, and no prior research has empirically examined blockchain technology's impact on consumer marketplaces. In collaboration with a leading Chinese technology company that operates the country's top online consumer grocery marketplace, we analyze the effects of its sellers' recent adoptions of blockchain tracing. We find that consumers respond to supply chain transparency by readily purchasing traced products, especially those that are handling-sensitive or sold by third-party sellers. For third-party sellers, information provided through blockchain tracing is fundamentally different than the sellers' own quality claims, leading to an up to 23.4% increase in average revenue for these sellers. By using structural estimation to analyze how consumers assess products, we highlight that consumers' prior experience with the product category shapes the sophistication and magnitude of their responses to tracing and the resulting welfare effects. Additionally, we show evidence that consumers deem blockchain-backed information to be markedly more reliable than sellers' own claims.

2 - The Environmental and Social Implications of Pay-What-You-Want Food Waste Supermarket

Luyi Gui, The Paul Merage School of Business, UC Irvine, Irvine, CA, United States, Xi Lin, Luyi Yang

Donating edible surplus food and making it available to food insecure communities is a preferred food recovery approach. An emerging type of donation outlet is pay-what-you-want food waste supermarkets where consumers can purchase edible salvaged food at a price that they are willing to pay as opposed to according to a posted price tag. In this paper, we analyze the impact of pay-what-you-want pricing for salvaged food on strategic consumer choices, and in turn, regular retailers' replenishment and donation decisions. Accordingly, we evaluate how pricing scheme choices of donation outlet could influence food waste generation and food insecurity reduction.

3 - Scope 3 Emission Disclosure in Supply Chains: Equity and Efficiency

Bin Hu, University of Texas at Dallas, Richardson, TX, United States, Ruize Ma, Yunke Mai

The United Nations has advocated the disclosure of Scope 3 (supply chain) carbon emissions. Two Scope 3 carbon emission assessments are commonly adopted: Cradle-to-Gate which includes emissions of all upstream firms, and Life-Cycle (or Cradle-to-Grave) which includes emissions of all upstream and downstream firms; in addition, one can conceptualize a Gate-to-Grave assessment which includes emissions of all downstream firms. We consider a serial supply chain where each firm discloses its carbon emissions subject to societal pressure, and study their emission-reduction and product-pricing decisions and the supply chain outcome under different carbon emission disclosure regulations in terms of carbon-reduction equity and efficiency. We find that Gate-to-Grave is consistently the most equitable assessment in terms of both carbon and profit reductions along the supply chain, whereas all assessments are virtually indistinguishable in carbon-reduction efficiency in terms of both consumer welfare and supply chain profit for the same total supply chain emission. These findings strongly advocate for the Gate-to-Grave Scope 3 emission assessment's formal definition, recognition and implementation.

4 - Innovator's Advantage In Competitive Technology Adoption: Enhancing Profit and Combating Food Waste via Supply Transparency

Chenghuai Li, Duke University, Durham, NC, United States, Bora Keskin, Jing-Sheng Song

Emerging digital technologies enable fresh produce retailers to trace and track the freshness conditions of their supplies in real time and communicate additional information, such as sustainable practice certifications, to consumers. These technologies offer retailers great opportunities to increase supply chain efficiency and reduce food waste. In a competitive retail market, we study whether an innovator of supply transparency technology has advantages over a follower. In a duopoly facing freshness-dependent demand, the adoption of transparency technology allows a retailer to observe the freshness state when placing orders and enhance consumer utility when selling fresh produce. We find that the innovator can benefit from capturing the demand spillover from competitor's stock-outs, thereby achieving greater profit growth compared to the follower. This spillover-capture effect is pronounced in the case of low retail margin as the non-adoper tends to under-order. Additionally, when technology creates consumer value through transparency, it can spur market growth, further amplifying the innovator's advantage through the market-expansion effect. Under certain conditions, the innovator can achieve a greater food waste reduction or more significant improvement in consumer welfare. We also find that the retail market with fewer competitors can have a higher propensity for industry-wide adoption, considering both the fixed and proportional costs of adoption. This finding suggests that mergers can reduce a grocery retail market's food waste and improve consumer demand fulfillment. We generalize our findings to the cases of freshness-based discounts and sequential technology adoption.

SB54

Summit - 446

Ethically and Environmentally Responsible Operations

Invited Session

MSOM: Sustainable Operations

Chair: Karthik Murali, Oregon State University, Corvallis, OR, United States

Co-Chair: Aditya Vedantam, State University of New York at Buffalo, Williamsville, NY, United States

1 - Renewable Energy Investments Under An Evolving Social Cost of Carbon

Thomas Palley, Indiana University, Bloomington, IN, United States, Owen Wu

The social cost of carbon (SCC) is an established measure for informing policymakers in making energy planning decisions. Uncertain and evolving SCC estimates present a challenge for the transition towards more renewable grids. We develop a model to examine how uncertainty and learning about SCC affects renewable energy investments, utilizing the relationship between atmospheric carbon levels and environmental and societal impacts. We consider a rational expectation equilibrium (REE) model in which individual firms make investment decisions based on their belief about the evolution of SCC, while the aggregate investment of all firms directly affects the atmospheric carbon accumulation and SCC. We analyze REE under various information structures to better understand how the speed of learning about components of SCC estimates can affect investments and emissions. We offer insights into how policymakers might improve decision making under uncertainty about the societal impact of carbon emissions.

2 - Renewable Energy Servicizing for Residential Homes in Developing Countries

Jaachinma Okafor, University of Tennessee, Knoxville, TN, United States, Russel Barton, Paolo Letizia, Gil Souza

Energy access remains a critical challenge in developing countries, characterized by the unavailability and unreliability of power supply systems. This paper explores the potential of adopting a servicizing approach to deliver renewable energy solutions to residential consumers in these regions. This proposed model involves a firm installing and maintaining a solar array-battery system in households in exchange for usage-based payments. By shifting ownership and maintenance responsibilities to service providers, households can gain access to clean and reliable solar energy without incurring upfront costs. The economic viability of this servicizing solution is examined by developing a stylized analytical model, considering factors such as investment costs, supply-demand variability, and consumer adoption. Our analysis derives the optimal charge per unit consumption of energy, the optimal solar generation capacity, and the optimal battery capacity, offering insights for firms aiming to provide sustainable energy solutions in developing nations. Notably, our results highlight the profitability of the servicizing model, especially in areas heavily reliant on gasoline generators for backup power. To further validate our findings, we conduct a trace simulation using actual solar irradiation and demand data from households in Nigeria. This simulation enhances the robustness of our results in a more realistic setting, showcasing the value of a shared system that efficiently serves multiple households.

3 - Empowering Supply Chain Energy Efficiency: Merits and Pitfalls of Buyer Collaboration

Jason Nguyen, Ivey Business School, Western University, London, ON, Canada

Environmental agencies in developing countries are increasingly partnering with large corporate buyers to enhance energy efficiency at local manufacturing facilities. We examine when such collaboration can improve social welfare and energy efficiency investments. We show buyer collaboration enhances energy efficiency investments, but it can harm the overall social welfare.

4 - Resale-as-a-Service: The Impact of Second-Hand Platforms on Fashion Goods Retailer

Robert Swinney, Duke University, Durham, NC, United States, Fernando Bernstein, Yuan Guo

We examine how a retailer selling fashion apparel should account for a second-hand market amid changing fashion trends and diverging customer tastes. We propose a customer decision model that incorporates both customer purchase decisions (i.e., purchasing a new product, a used product, or no product) as well as their selling decisions (i.e., selling a used product) over time. Customers have heterogeneous tastes for product design and each product released by the retailer has random popularity in the market. We discuss how uncertainty in customer taste over time can affect the retailer’s design and production decisions when the resale platform operates independently from the brand or retailer. Furthermore, we investigate the implications of collaborating with such a platform and embracing Resale-as-a-Service (RaaS) for different types of retailer’s design and production decisions. We examine the potential benefit of RaaS for both parties involved and discuss its role in the supply chain.

5 - Survival Analysis for Managing Animal Rescue

Elena Katok, The University of Texas at Dallas, Richardson, TX, United States, Qiuxia Chen, Ernan Haruvy

Municipal shelters must euthanize numerous animals (primarily dogs) every day due to a shortage of space. The choice of which dogs to euthanize is based on two policy decisions: the length of time the dog is given at the shelter, and the degree to which the shelter capacity can be expanded. The more the shelter is over capacity, the more stressful it is for the dogs who are there, due to noise, overcrowding, and a lower quality of care, therefore, shelter managers have to consider the trade-off between the higher probability of a dog being adopted when it stays longer, and the lower probability of a dog being adopted due to over-crowded conditions. Using data from a large municipal shelter in Dallas, we derive a method for determining the optimal policies using dual-objective survival analysis. This analysis allows us to qualitatively estimate the number of lives saved compared to the current practice. Our work can be generalized to the problem of when to liquidate excess inventory.

SB55

Summit - 447

Humanitarian and Disaster Operations Management II

Invited Session

Public Sector OR

Chair: Andrew Arnette, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Christopher Zobel, Virginia Tech, Blacksburg, VA, United States

1 - Media Coverage of Disasters and Its Implications for Precautionary Purchases

Trilce Encarnacion, University of Missouri- St. Louis, Saint Louis, MO, United States, Johanna Amaya

In preparation for disasters, households purchase large quantities of supplies to ensure they have enough resources during emergencies. This phenomenon is commonly referred to as "Panic Buying." These purchases can severely disrupt supply chains by creating sudden, extreme demand spikes that lead to stock shortages, logistical bottlenecks, and price inflation, making it difficult for retailers to maintain steady supply and for consumers to access essential goods. Understanding the influencers of these precautionary purchases is crucial for developing strategies to mitigate its impact on supply chains and ensure equitable access to essential supplies during emergencies. To explore the dynamics of this behavior, we analyze the relationship between media influence and consumer purchasing patterns by integrating two key data sources. First, we use news reports about hurricanes that have impacted the United States, which provide a timeline and intensity of media coverage before and during these events. Second, we utilize retail scanner data that records sales of emergency supplies such as bottled water, batteries, and non-perishable foods. Leveraging Theory of Planned Behavior and Media Influence Theory as theoretical basis, we develop econometric models to quantify how media coverage triggers precautionary buying behaviors among consumers.

2 - A Resource Allocation Model to Mitigate the Risk of Disruptive Events in Homeland Security

Parastoo Akbari, Iowa State University, Ames, IA, United States, Cameron MacKenzie

A disruptive event in homeland security, whether it is a natural, technological, or human-caused event, that impacts the homeland can have severe consequences. The Department of Homeland Security and state governments in homeland security and emergency management are responsible for allocating resources and funding projects that reduce the risk of these disruptive events. This research proposes a resource allocation model that enables a decision maker to choose among a set of projects that mitigate the risk of disruptive events in a geographic area while satisfying a budget constraint. The risk of a disruptive event is measured as the product of the probability of the event and the multidimensional consequences if the event occurs. We apply the resource allocation model to the state of Iowa and identify 16 different hazards that threaten the state including floods, tornadoes, winter storms, bridge failure, a cyberattack, and an improvised explosive device. Based on a variety of information sources, we assess an annual occurrence probability for each event and 6 consequences (injuries, fatalities, property damage, crop damage, number of customers without power, and number of businesses closed). We estimate the costs and effectiveness of 52 projects described in the State of Iowa’s homeland security planning documents. We solve the resulting optimization problem to select the projects that best mitigate the risk of these 16 different hazards. This resource allocation model can improve the selection of projects and assign homeland security resources more effectively based on risk.

3 - Optimizing Humanitarian Aid Delivery in Disaster Zones: a Mathematical Model for Collaborative Truck and Drone Logistics

Ramin Talebi Khameneh, Stevens Institute of Technology, Hoboken, NJ, United States, Nafiseh Ghorbani-Renani, Jose Emmanuel Ramirez-Marquez

In post-disaster urban areas, humanitarian aid operations are not only essential but also complex because of dense populations and damaged infrastructure. Despite the common logistics, humanitarian logistics must manage a variety of critical objectives beyond minimization of cost, ensuring quick and equitable distribution of aid to affected populations. These challenges highlight the importance of strategies that optimize the use of diverse technologies and modalities. In spite of significant advances in transportation technologies, there is still a considerable gap in coordinating multiple transport modes to optimize aid delivery time and efficiency in urban areas affected by disasters, such as trucks and drones. In this study, we propose a mathematical model for solving the cooperative trucks and drones humanitarian relief delivery problem where each drone is associated with a particular truck, and both are dispatched from a central depot simultaneously to serve aid recipients. The implementation of this model aims to significantly reduce total delivery time, enhancing the speed and precision of aid deployment with key innovations, including optimizing route paths for both trucks and drones within operational constraints like battery life and payload capacities. From a practical perspective, the case study of urban flooding illustrated the potential to dramatically increase disaster preparedness and response strategies, thereby increasing the resilience of urban environments.

4 - Optimizing Spatial Prepositioning of Food Supplies: A Covering Model for Disaster Resilience in Germany

Sonja Rosenberg, Karlsruhe Institute of Technology, Karlsruhe, Germany, Katharina Eberhardt, Frank Schultmann

Public food stockholding is critical in facilitating recovery from disruptions like natural or human-made disasters, enhancing a nation's resilience. However, many challenges exist in planning food stockpiling, including adequate pre-positioning of supplies. This challenge is even more significant if it is conducted for disasters with a low probability of occurrence but a high degree of vulnerability for a nation. Disasters, such as pandemics or severe power blackouts, may be long-lasting and nationwide. Some nations, like Germany, prepare for these circumstances by stockpiling durable goods, e.g., wheat, in defined warehouses over long periods. Because only limited aid can be expected from surrounding regions in case of large-spatial disasters, the pre-positioning of durable goods must ensure an adequate covering of the complete area. Therefore, we propose a covering model that maximizes the demand coverage of a population during a food crisis by optimizing the warehouse locations, their capacities, and the individual covering zones under a limited public budget. Trade-offs exist between economies of scale for warehouse sizing and efficient use of transport infrastructure when distributing the supplies. The model is applied to Germany as a case study, investigating the degree to which spatial vulnerability indicators of the population affect the model results.

5 - Does Gender Shape Organizers' Charitable Crowdfunding Success?

Margaret Traeger, University of Notre Dame, Notre Dame, IN, United States

Do donors on a charitable crowdfunding platform give more money to women or men campaign organizers? Examining the role of organizers' gender in this context allows us to address underexplored theoretical questions about how gender shapes interactions in a context with clear feminine and masculine elements. Social psychological theories of gender predict that men's presumed higher status will advantage them in most situations, except communal contexts in which women are expected to excel. Because the charitable crowdfunding context combines elements of both caring-oriented and achievement-oriented contexts, it is possible that either women or men could perform better. Despite the strong communal content of charitable crowdfunding, we find men organizers raise substantially more money than women. Gender differences in campaign design do not explain the disparity. There are additional inequities in how donors respond to the type of language women and men use in their campaigns. Agentic words increase donations for women and men, but agentic language is associated with larger donations to men than women. Strikingly, communal language increases donations to men, but decreases donations to women's campaigns. Our results suggest donors hold unconscious gender biases about men's greater status that drive donors to lend more money to men's campaigns than women's campaigns.

SB56

Summit - 448

Substance Use Disorders and Other Stigmatized Medical Needs

Panel Session

Women in OR/MS (WORMS)

Co-Chair: Wilkistar Otieno, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Co-Chair: Di Nguyen, University College Dublin, Dublin, Ireland

1 - Moderator Panelist

Di Nguyen, University College Dublin, Dublin, Ireland

2 - Moderator Panelist

Wilkistar Otieno, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

3 - Panelist

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

4 - Panelist

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

5 - Research in Opioid Mitigation

Mary Beth Kurz, Clemson University, Clemson, SC, United States

6 - Panelist

Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

7 - Panelist

Veronica White, FAMU-FSU College of Engineering, Tallahassee, FL, United States

SB57

Summit - Terrace Suite 1

Innovative Applications of Optimization Techniques in Healthcare

Invited Session

Health Applications Society

Chair: Nasrin Yousefi, Queen's University, Toronto, ON, Canada

1 - Patient-to-Bed Placement Optimization: Improving Coordination for a Hospital Command Center

Arlen Dean, University of Michigan, Ann Arbor, MI, United States, Mark Van Oyen, Mohammad Zhalechian

We present our work with a large hospital to improve coordination among bed managers for patient-to-bed placements. Hospital bed managers are responsible for managing patient flow, a critical aspect of hospital operations and patient care. Allocating beds can be complex and time-consuming due to the shared resource nature of beds and the varying care needs of patients. To address these challenges, we developed and implemented an algorithm based on an optimization model that uses operational practices and real-time patient and system information. We examine the effects of our algorithm on our partner hospital's operations post-implementation using an empirical analysis based on several key outcome metrics related to care quality.

2 - Screening Process Information Matters More than Clinical Information: a Randomized Controlled Trial of Information Sharing on Cancer Screening Decision

Sarah Yini Gao, Singapore Management University, Singapore, Singapore, Xiaodong Wang, Zhichao Zheng, Yongjian Zhu

Policymakers often rely on information sharing for decision-making, especially in healthcare. The effectiveness of sharing the right information is pivotal for influencing patient behavior, particularly in encouraging cancer screening uptake, yet the literature lacks consensus regarding the most effective elements of information to share. This study investigates the specific types of information that should be disseminated to encourage cancer screening uptake.

By analyzing data from a randomized controlled trial focused on colorectal cancer screening, our findings reveal that sharing relevant information significantly improves participants' intention to take up screening, independent of their initial level of knowledge before the intervention. Specifically, sharing information on the screening process improves the screening intention significantly, whereas the effect of sharing clinical information about cancer is insignificant. Furthermore, sharing information that is already familiar to participants proves to be just as effective as introducing new information to enhance the screening intention. We also identify heterogeneous responses from different subpopulations to different types of information content.

Our findings suggest that identifying knowledge gaps before sharing information might be less crucial than previously thought, as reinforcing known information also proves beneficial. For public health campaigns, emphasizing the screening process and tailoring information to diverse subpopulation reactions are recommended to optimize effectiveness.

3 - Quantifying the Benefits of Customized Vaccination Strategies: a Network-Based Optimization Approach

Su Li, Texas A&M University, College Station, TX, United States, Hrayer Aprahamian

We study the problem of designing vaccine distribution strategies that maximally mitigate the negative impact of an infectious disease outbreak. This is achieved through a multiperiod optimization-based framework that embeds important subject-specific risk and contact information into the decision-making process. By analyzing the structure of the resulting optimization problem, we identify key structural properties which we use to construct a globally convergent solution scheme (suitable for smaller problem instances) and two, more scalable, heuristic schemes. We demonstrate the benefits of the considered framework through a case study on COVID-19 in Texas. Our results highlight the importance of considering risk and contact information as doing so substantially reduces the total expected number of fatalities over conventional compartmental-based approaches. These findings indicate that customization can have a significant benefit, particularly for community-scale planning.

4 - Uncertainty Quantification in Inverse Optimization

Nasrin Yousefi, Queen's University, Toronto, ON, Canada, Timothy Chan, Nathan Sandholtz

Decisions often exhibit "noise" around theoretical optimums in applied optimization tasks. Noise may arise from measurement, model, or human decision-making errors. The presence of noise adds significant complexity when considering the inverse optimization problem, which is the problem of inferring unknown parameters of an optimization model such that a collection of observed decisions is rendered optimal for the inferred model. While several works have proposed solutions for the noisy inverse optimization problem in terms of point estimation, they have not directly addressed the uncertainty in their estimates. Although uncertainty is inherent in the context of noisy inverse optimization, the quantification of this uncertainty has been treated very little in the literature. Our work is the first to address this topic directly. We study optimization problems with linear objective functions and convex feasible regions where the objective coefficient vectors are unknown. We propose a parametric approach to estimate the unknown coefficient vectors and construct credible regions around the estimates.

5 - Feasibility and Optimization of Prehospital Critical Care Team Response to Major Trauma in the Greater Toronto Area

Rachel Stephenson, University of Toronto, Toronto, ON, Canada

In the United Kingdom and Europe, it is common for physician-led prehospital critical care teams (PHCCTs) to respond to major traumas at the scene of the incident. PHCCTs provide interventions outside of the competencies of standard paramedics which, while required infrequently relative to other standard paramedic-administered interventions, can be lifesaving for major trauma patients. We aim to assess the feasibility of PHCCT response to major trauma in the Greater Toronto Area (GTA). We use geospatial modeling and maximal covering location models to determine the optimal base locations for PHCCT teams in the GTA to respond to major traumas under different expert-informed emergency medical services response configurations. We evaluate our proposed solutions using discrete event simulation considering both total coverage as well as the equity and fairness implications of allocating a scarce health resource.

SB58

Summit - Terrace Suite 2

Bonder Scholars Session

Award Session

Health Applications Society

Chair: Yuming Sun, Georgia Institute of Technology, ATLANTA, GA, United States

1 - Online Learning with Survival Data

Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States, Hamsa Bastani, John Silberholz

This project adapts online learning techniques to scenarios with time-to-event data, where there is a delay between choosing an arm and observing feedback that is endogenous to the quality of the arm. We posit a multi-armed bandit algorithm that uses a cox-proportional hazards estimator. We theoretically analyze and prove guarantees on the regret under this algorithm. Lastly, we examine its performance on a dataset of metastatic breast cancer clinical trials, and compare it to that of other adaptive allocation schemes.

2 - Diabetes Treatment Planning Using Nonstationary Combinatorial Multi-Armed Bandits

Katherine Adams, University of Wisconsin-Madison, Madison, WI, United States, Justin Boutilier, Sarang Deo, Yonatan Mintz

Unhealthy diets and sedentary lifestyles are key drivers of the global diabetes pandemic. Research shows office workers are particularly at risk of developing diabetes. The chronic nature of diabetes burdens firms with increased labor costs due to reduced productivity, absenteeism, and turnover. Partnering with Community Health Worker (CHW) organizations provides cost-effective solutions for early detection and management of diabetes symptoms, driving a healthier, more productive workforce. Building efficient CHW visit plans require modeling tradeoffs between screenings, provision of care to the enrolled, and visits to gather information on each patients' disease status. To address this problem, we propose a nonstationary combinatorial multi-armed bandit framework that allows providers to build personalized CHW visit plans to screen and manage patients with diabetes under uncertainty.

3 - Quantifying the Impact of Vaccinating Under-Immunized Groups in Polio Outbreak Responses: a Model-Based Study

Yuming Sun, Georgia Tech ISyE, Atlanta, GA, United States, Hongyu Xue, Pinar Keskinocak, Lauren Steimle

Polio, an infectious disease that causes paralysis, remains a global health concern, especially with the emergence of circulating vaccine-derived poliovirus (cVDPV) and recurring outbreaks in areas with cohorts of under-immunized individuals. This study systematically assessed how the allocation of vaccines during a polio outbreak response might impact outcomes. Adapting a compartmental simulation model, we projected poliovirus transmission from 2024 to 2026 under different levels of vaccination campaign coverage (i.e., the proportion of the target population planned to be vaccinated), vaccine allocation schemes (e.g., across different immunity groups), and vaccination campaign delays. Results highlighted that compared to other allocation schemes, priority allocation of vaccines to under-immunized immunity groups (i) significantly reduced the case burden, even with lower coverage and longer delay; (ii) achieved die-out of transmission with two rounds of vaccination if the delay was short (<= 3 weeks) and coverage was high (>= 70%).

4 - Interpretable and Equitable Population Guidelines

Sun Ju Lee, Georgia Institute of Technology, Atlanta, GA, United States, Gian Garcia

Evidence-based guidelines play an important role in informing how chronic diseases should be managed, as these recommendations are widely disseminated and widely implemented. However, they are one-size-fits-most, failing to account for patient-to-patient differences. Personalized medicine has shown significant potential to improve health outcomes over clinical practice guidelines. However, the implementation of personalized medicine may be challenging, resulting in unwanted practice variation and suboptimal patient care. To find the right balance between personalized medicine and clinical guidelines, we propose the treatment guideline design problem. We develop a framework to design optimal treatment guidelines for a population of patients which are each modeled according to their own Markov Decision Process parameters. We propose several exact and heuristic methods to solve the problem, including a mixed-integer linear program formulation and a branch-and-bound clustering algorithm. Our results on a hypertension treatment planning case study demonstrate that with only a small number of subgroups, we can develop clinical practice guidelines which perform almost as well as personalized treatment policies.

SB59

Summit - Ballroom 1

Where Were you When you First Heard about OR -- and What About the Future of the Profession

Panel Session

Committee’s Choice

Co-Chair: Daniel Bienstock, Columbia University, New York, NY, United States

1 - Moderator Panelist

Daniel Bienstock, Columbia University, New York, NY, United States

2 - Panelist

RICHARD O'NEILL, none, Silver Spring, MD, United States

TBD

3 - Panelist

Garrett van Ryzin, Amazon, Brooklyn, NY, United States

4 - Panelist

Irem Sengul Orgut, University of Alabama, Tuscaloosa, AL, United States

5 - Panelist

Alexandra Newman, Colorado School of Mines, Golden, CO, United States

6 - Panelist

Andres Gomez, University of Southern California, Los Angeles, CA, United States

SB60

Summit - Ballroom 2

Order Fulfillment Optimization within 4 walls of Amazon Warehouses

Invited Session

The Practice Section of INFORMS

Chair: Ignacio Erazo, Amazon, Bellevue, WA, 98004, United States

1 - Optimizing Storage Across Bin Types in Traditional Fulfillment Centers

Kevin Bunn, Amazon, Bellevue, WA, United States

In this presentation, we look at a traditional fulfillment centers (FCs), where items are stowed in different types of bins based on the shapes and sizes of those items. In many cases, items can be stowed in multiple bin types. The goal of this research is to propose a strategy that considers the current fullness levels of different bin types, and makes a recommendation for which bin type to stow items in order to optimize the FC's overall stow and pick rates. In order to do so, we model how stow and pick rates are impacted by fullness levels for each bin type. We then estimate how many items we plan to stow and pick per day, and the distribution of bin types where items can be stowed. From there we can set an optimal fullness level across each bin type which minimizes the amount of time it would take to stow and pick. We also plan consolidation, which is the process of moving inventory between bins, to identify situations where a consolidation plan can improve fullness levels.

2 - Manual Pick Scheduling Optimization in Traditional Fulfillment Centers

venkatachalam avadiappan, Amazon, Bellevue, WA, United States, Kay Zheng

In the Amazon traditional fulfillment centers (FC), the pickers move to the inventory stored in bins and manually pick the assigned demands, place them in carts and send them downstream for consolidation and packaging. The picking services play a key role in the order fulfillment process in a FC and significantly impacts the FC speed and throughput. They determine which inventory to use to fulfill a demand, at what time, and by which picker so that it can be shipped out on time to meet the customer promise. Currently, the demand to inventory binding decisions are pre-determined (i.e., early binding), while picker assignment and picker path routing are based on greedy heuristics, resulting in long picker travel distances and low pick throughput. We propose a novel manual pick scheduling (MaPS) optimization model which improves the demand to inventory binding (i.e., late binding) and picker path routing and solves within reasonable times, thereby reducing distance traveled by pickers and enhancing pick throughput. MaPS is decomposed into a two-stage sequential optimization to address scalability challenges, wherein, in the first stage, selection of demand from backlog, demand to inventory binding, picker assignments are determined, while in the second stage, the picker routing decisions are made. Preliminary results suggest that MaPS significantly improves the average expected travel time between picks compared to the heuristic.

3 - Reinventing Robotic-Assisted Picking Algorithms in Amazon

Kay Zheng, Amazon, Bellevue, WA, United States, Ignacio Erazo

Fulfilling orders at the scale and speed that Amazon does is a challenging task that requires optimized processes throughout an order's life cycle. Amazon Robotics (AR) Fulfillment Centers (FC) have proved to be more efficient and ergonomics-friendly than its traditional counterparts, empowered by the robotic technology that enables stationary and continuous picking from inventory pods that are brought to pickers by robotic drives. AR FCs face different challenges than traditional FCs and requires innovative ways to solve for a pick execution plan quickly while accounting for associate safety. Today’s pick planning in AR FCs is driven by a heuristic that can quickly create short-term pick execution plans to meet upstream targets; however, its greedy nature has shown to be very limited in leveraging all the data and optimizing for multiple objectives. In this talk, we present a new optimization-based, scalable and flexible framework for AR pick planning and execution that coherently optimizes for upstream targets adherence, pick productivity and associate safety.

4 - Prioritizing Demand Picking with Multiple Objectives

Elcin Cetinkaya, Amazon.com, Seattle, WA, United States, Prem Viswanathan, Kevin Bunn

A specific service at Amazon Order Fulfillment Systems prioritizes shipments within a Fulfillment Center driven by a multi-objective optimization model. The service is based on a model predictive control framework whose output is in the form of a multi-hour plan for picking that's refreshed multiple times during the day. The plan is then used to construct a set of nested shipment groups from which the picking service has the flexibility to select a subset of work. In this talk, we will present the multi-objective optimization model and nested shipment grouping framework.

5 - Scaling Up Pick Planning in Amazon

Ignacio Erazo, Amazon, Bellevue, WA, United States, Kay Zheng

Order picking is a vital component in an order's fulfillment journey, and the creation of picking plans is subject to multiple business requirements and operational targets that are updated every 2.5 minutes by upstream systems. Pick planning algorithms for Amazon Robotics (AR) Fulfillment Centers (FC) must be able to react to these moving operational targets as soon as possible, hence the need to create new pick execution plans within 2.5 minutes. Because of this challenge, Amazon created a fast heuristic that creates feasible plans in a somewhat greedy manner. While the heuristic has proven to be reliable, we have observed degradation of pick productivity since this algorithm's gradual network roll-out. Moreover, the current methods have difficulty creating solutions of high quality when considering the new use cases in AR FCs. We propose an optimization-based algorithm that decomposes the original pick planning problem with multiple millions of integer variables into at least two phases, allowing to generate picking plans that meet the business constraints and improve picking efficiency all while scaling to the new use cases and solving within the time constraints. Experiments show that our new methodology improves significantly the quality of the picking plans, and we also demonstrate that it is not needed to solve the new models to optimality to obtain such improvement. Furthermore, by using our methodology offline to train a machine learning model, we are able to reduce by 18% the size of our problem, without loss of optimality.

SB61

Summit - Ballroom 3

Why I Love Optimization

Panel Session

Committee's Choice

Co-Chair: Amy Cohn, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

1 - Moderator Panelist

Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

2 - Panelist

Benjamin Grimmer, Johns Hopkins University, Baltimore, MD, United States

3 - Panelist

Siqian Shen, U of Michigan/NSF, Ann Arbor/Alexandria, MI, United States

4 - Panelist

Dan Iancu, Stanford University, Stanford, CA, United States

5 - Moderator Panelist

Amy Cohn, University of Michigan, Ann Arbor, MI, United States

SB62

Summit - Signature Room

Humanitarian Operations and Earmarked Funding

Invited Session

TutORial

Chair: Harish Krishnan, University of British Columbia, Vancouver, BC, Canada

1 - Humanitarian Operations and Earmarked Funding

Alfonso Pedraza-Martinez, University of Notre Dame, Notre Dame, IN, United States

Earmarked funding, also referred to as restricted funding, is one of the main characteristics of humanitarian operations. Earmarked funding can be defined as the donors' contributions to humanitarian organizations to be used for a specific purpose. This is in contrast to flexible funding, which can be used for any purpose. This tutorial introduces the trade-off between total donations and operational performance due to humanitarian earmarking. The tutorial explains why allowing donors to earmark their contributions helps organizations increase fundraising effectiveness. It also explains why earmarking hurts organizations' per-dollar performance. Because humanitarian organizations' utility increases in total donations and per-dollar (or any other currency) performance, the best fundraising strategy for the organizations, collecting earmarked or flexible funds, is not apparent. Moreover, this tutorial argues that ear marked funding is here to stay and discusses models that reduce the negative effect of earmarked funding. Then, it proposes that a thoughtful mix of earmarked and flexible donations may be the best way for humanitarian organizations to fund their operations. The balance between earmarked and flexible funds depends on reducing the negative operational consequences of earmarked funds. The tutorial concludes by identifying areas for future research on humanitarian operations and earmarked funding.

SB63

Regency - 601

Economics of Information Systems

Invited Session

Information Systems

Chair: Stephanie Lee, University of Washington, Seattle, WA, United States

1 - Pseudonymity, Anonymity and Online Discussion Community

Jiaying Deng, Fordham University, New York, NY, United States, Lin Hao, Yingda Lu, Ali Tafti

Users' active engagement is strongly correlated with the success of online discussion community and identity plays an important role in shifting individuals' participation. Previous literature often entails the comparison between anonymity and offline identity in individual engagement on social media platforms. However, the effects of pseudonymity on users' content-generation behaviors are nonetheless under-studied. While both pseudonymity and anonymity involve concealing one's offline identity, pseudonymity involves the use of fictitious online identity (such as usernames) whereas anonymity entails the complete absence of personally identifiable information. In this study, we examine the impact of pseudonyms on users' engagement within online discussion forums through a unique natural experiment where our treatment forum removed all usernames. Our findings reveal distinct user responses to pseudonymous settings compared with previous literature on anonymity. Notably, individuals are less likely to ask questions or provide answers after concealing their pseudonyms. In addition, the negative impact is more significant for users who are more centrally embedded in a network. Nevertheless, contrary to previous research suggesting that anonymity contributes to Internet trolling, we find no evidence of an increase in trolling behavior when pseudonyms become anonymous. Our study contributes to the literature on user content generation and the development of identity policies in online discussion community.

2 - News in the Dark: Effects of Facebook's Australian News Ban on News Consumption

Ying Gu, Univesity of Washington, Seattle, WA, United States, Stephanie Lee, Yong Tan

In the contemporary news consumption landscape, social media platforms have emerged as key sources of news consumption, and there have been ongoing discussions on the relationship between social media platforms and news publishers. In 2021, the Australian News Media Bargaining Code was proposed to mandate platforms to pay news publishers for their content. In response, Facebook removed news content from the platform in Australia from February 18th to 26th, 2021. This study examines how Facebook's decision to restrict news sharing in Australia affected news consumption in Australia. Using a difference-in-differences approach, along with the propensity score matching, we find a decline in news page views in Australia post-ban, indicating significant changes in news engagement patterns, particularly among casual readers who predominantly relied on Facebook for news dissemination. The paper extends beyond merely documenting this decline, looking into the broader implications of such platform policies on digital news consumption landscapes. It uncovers shifts in device usage, with a notable transition from mobile to desktop engagement, and finds resilience in user behavior, evidenced by increased traffic from search engines post-ban. These findings offer a valuable empirical perspective on the dynamics shaping digital news engagement, providing key takeaways for policymakers, news organizations, and digital platforms navigating the complex landscape of news dissemination in the digital age.

3 - Netflix and Chilling Effect? the Impact of Streaming Video Licenses on Paid Digital Downloads

Shahryar Doosti, Chapman University, Orange, CA, United States, Brett Danaher

Digital distribution of entertainment products has led to new opportunities for firms to extract revenues through a variety of different channels but also new challenges in determining how to best manage those channels, given the potential for complementarities and cannibalization. In particular, television networks and movie studios are faced with the challenge of negotiating fixed-fee licensing deals for their content on streaming services such as Netflix without knowing what effect such deals will have on the demand through other channels. In this research, we exploit the natural experiment of Netflix's launch in Australia to determine the causal effect of licensing a title to this service on paid digital downloads of that content. We find that licensing TV shows to Netflix caused digital sales of those shows to decrease by an average of 31% relative to shows that were not licensed, while licensing films caused a 52% decrease. We also find evidence that films continue to experience a depressive effect on digital downloads even after they are removed from Netflix. Our results shed light on the effect that consumption through a subscription bundle has on a la carte purchases, and they also have practical implications for content owners negotiating deals for their content with streaming video services.

4 - From Bytes to Bites: Enhancing Personalized Nutrition Interventions for Diabetes Management Through Causal Inference with Double ML and Large Language Models

Rebekah Baik, University of Washington, SEATTLE, WA, United States, Leonard Boussioux

Personalized nutrition plays a crucial role in the prevention and management of diabetes. However, understanding the causal relationships between dietary factors and blood glucose levels remains a complex challenge. This research proposes a novel causal inference framework that leverages the power of large language models (LLMs) and double machine learning (DML) to integrate multimodal data and uncover complex patterns in the diet-diabetes relationship. By generating rich embeddings from both structured and unstructured data, the proposed approach enhances the input features for the DML model, which employs tree-based models for treatment and outcome estimation. This integration allows for the identification of the causal impact of specific dietary interventions while controlling for confounding variables. A key focus of this framework is to improve the interpretability of treatment effects, facilitating the development of targeted dietary recommendations and optimizing healthcare resource allocation for diabetes prevention and management. This research aims to bridge the gap between complex causal inference techniques and their practical application in healthcare decision-making by providing a more understandable and explainable model. The insights gained from this framework can lead to more efficient and cost-effective personalized prescriptive nutrition strategies, ultimately benefiting individual health outcomes and healthcare systems.

SB64

Regency - 602

Social Media Analytics Student Paper Contest

Award Session

Social Media Analytics

Chair: Lingshu Hu, Washington and Lee University, Lexington, VA, United States

1 - Measuring the Impact of Unobserved Social Networks on Group-buying Decisions under Privacy Concerns

Jingwen Zhang, University of Washington, Seattle, WA, United States, Jin Liu, Yongjun Li, Yong Tan

Leveraging the viral power of social networks, group buying has rapidly emerged as a promising e-commerce model with substantial market potential. Companies have been increasingly integrating group-buying campaigns into social media platforms to further accelerate demand growth. However, tapping into these expansive social networks heightens users' privacy concerns - not just around corporate data practices but also regarding the uncontrolled exposure risks as personal information propagates across widening social circles. The inability to observe the underlying peer connections in most group-buying datasets has posed significant challenges for researchers seeking to identify social network effects on group-buying behaviors, especially social network effects of privacy preferences. We address this obstacle through applying an econometric technique tailored for social network effects with unobserved networks into the group buying context. Leveraging a novel group-buying dataset on a Chinese social media platform, we measure the impacts of social networks on consumers' sign up behaviors in group buying while the social network is unobserved. We find that (1) An individual consumer is more likely to participate in group buying when close friends have signed up, indicating a positive peer effect. (2) A privacy-sensitive individual consumer is less likely to participate in group buying themselves. Yet when close friends with high privacy sensitivity sign up, the individual consumer would be more likely to participate in group buying, indicating a surprisingly positive contextual effect of privacy preference. Our findings provide important implications for platforms to design group-buying mechanisms.

2 - Content-Centered or Creator-Centered? An Economic Analysis of Fan Loyalty on User-Generated Content Platforms

Meilin Gu, Tsinghua University, Beijing, China, People's Republic of, Zizheng Liu, Dengpan Liu

The loyalty of fan communities to content creators significantly impacts stakeholders on user-generated content (UGC) platforms. Notably, major UGC platforms exhibit varying stances on creators' cultivation of fan loyalty. For instance, Douyin (the Chinese version of TikTok) hinders creators from building robust fan followings, while Kuaishou (another major UGC platform in China) takes active measures to encourage and support creators in cultivating fan loyalty on its platform. This stark contrast has sparked an ongoing debate about the role of UGC platforms in the cultivation of fan loyalty. In this paper, using a game-theoretic model, we explore the optimal strategies embraced by platforms in managing creators' cultivation of fan loyalty and their implications for key stakeholders. Our findings challenge the conventional wisdom that suggests UGC platforms should always facilitate creators in cultivating fan loyalty, as such assistance may sometimes hurt the platforms. Furthermore, our study generates meaningful implications by providing guidelines for UGC platforms to strategically manage creators' cultivation of fan loyalty and for creators to balance their efforts between cultivating fan loyalty and enhancing content quality.

3 - Shaping Opinions in Social Networks with Shadow Banning

Yen-Shao Chen, Yale University, New Haven, CT, United States, Tauhid Zaman

The proliferation of harmful content and misinformation on social networks necessitates content moderation policies to maintain platform health. One such policy is shadow banning, which limits content visibility. The danger of shadow banning is that it can be misused by social media platforms to manipulate opinions. Here we present an optimization based approach to shadow banning that can shape opinions into a desired distribution and scale to large networks. Simulations on real network topologies show that our shadow banning policies can shift opinions and increase or decrease opinion polarization. We find that if one shadow bans with the aim of shifting opinions in a certain direction, the resulting shadow banning policy can appear neutral. This shows the potential for social media platforms to misuse shadow banning without being detected. Our results demonstrate the power and danger of shadow banning for opinion manipulation in social networks.

4 - The Interplay of Opinions and Behavior in Social Trading: Early Cryptocurrency Adoption Dynamics

Ye Liu, University of Washington, Seattle, WA, United States, Mingwen Yang, Matthias Pelster, Yong Tan

Social trading platforms allow individual investors to access peer opinions to make their investment decisions while transparently observing their peers' trading behaviors. We exploit a unique dataset from a large social trading platform and study how the interplay of opinions and behaviors within social networks influences an individual's decision to adopt cryptocurrency. We focus on cryptocurrency adoption, as cryptocurrencies have gained increasing attention over the last few years. Due to the lack of standardized information, social media has become a main information channel for individuals to decide whether to trade cryptocurrencies. This reliance on social media presents a unique challenge: The vast and unregulated nature of social media can often lead to disseminating information of questionable credibility. Using investment opinions and actual trading behaviors, we find a pronounced social learning effect in cryptocurrency adoption. Individuals are more inclined to adopt Bitcoin when their network peers do, especially when these peers' trading actions are consistent with their expressed opinions. In times of high uncertainty, investors tend to rely more on positive opinions, regardless of others' actual trading actions. In addition, high-income and older investors are more susceptible to opinions from their neighbors. Textual analysis indicates that Bitcoin posts are longer and more forward-looking compared to non-Bitcoin posts. We rule out alternative mechanisms such as awareness or "Keeping up with the Joneses."

SB65

Regency - 603

Pedagogical Design and Innovations for Analytics Education

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Tiffany Bayley, Ivey Business School, London, ON, Canada

1 - Enhancing Systems Thinking Skills in Healthcare Management Through Experiential Learning with &Ldquo;Friday Night in the Er&Rdquo;

Natallia Gray, Iowa State University, Ames, IA, United States

Healthcare professionals today are encouraged to build their skills in teamwork, systems thinking, and interprofessional collaboration to enhance patient safety and the quality of care. However, the avenues for effectively learning and practicing these skills are often scarce. Additionally, health educators face the challenge of finding effective methods to teach systems thinking.

To address this need, we integrated "Friday Night in the ER," an experiential learning tabletop game, into an existing healthcare management course. This game provides a hands-on approach to teaching systems thinking by immersing participants in the role of a hospital department manager during a condensed 24-hour simulation. Through this activity, participants learn to apply core principles of systems thinking to optimize system-wide performance, and receive scores based on quality of care, cost, and efficiency.

Additionally, we facilitate an industry workshop where healthcare administrative professionals engage in the same game. By comparing performance metrics between healthcare management students and seasoned professionals, we uncover valuable insights into potential gaps in systems thinking skills across the two groups.

To measure the effectiveness of the game in teaching systems thinking, both groups of participants are asked to complete a Likert scale survey. This survey evaluates the extent to which the activity contributes to achieving the game's learning objectives.

2 - The Feature-Based Newsvendor Model: The Fusion of Predictive and Prescriptive Approaches

Vahid Roshanaei, Rotman School of Management, University of Toronto, Toronto, ON, Canada, AKRAM KHALEGHEI

In this talk, we will present the newly designed and implemented Python-based lecture on the feature-based newsvendor problem (F-NVP), which has been integrated into the curriculum of the "Management Analytics" course at the Rotman School of Management, University of Toronto. Unlike the traditional newsvendor model, which concentrates solely on individual products, F-NVP integrates features from analogous products to provide a more nuanced comprehension of demand dynamics, a particularly beneficial approach for new product launches where historical data may be scarce or absent. By quantifying the average contribution of each feature in generating demands for similar products, F-NVP helps decision-makers better predict the expected demand for the product about to be released, based on its features. Our talk covers aspects such as the (i) non-linear formulation of the F-NVP, (ii) techniques for its linearization, and (iii) the application of the traditional two-step predict-then-optimize approach and the new single-step machine-learning approaches to address "big data" challenges. We will illustrate the superiority of this single-step approach over the traditional two-step predict-then-optimize method using a women's jeans product from Zara as an example. This lecture provides students with insights into inventory optimization strategies. By bridging theory with hands-on Python coding, students gain a deeper understanding of inventory management principles and their applications in today's data-driven business landscape.

3 - Strategic Course Sequencing in Business Education: a Mixed-Method Inquiry

Tiffany Bayley, Ivey Business School, London, ON, Canada, Kyle Maclean, Mazi Raz

Business school curricula traditionally encompass courses on analytics, strategy, marketing, organization behaviour, operations management, finance, accounting, and more. These topics are covered over the duration of 3- or 4-year program. However, the order in which these courses are introduced varies across institutions. Motivated by discussions during our institution's recent quality assurance process, we explore 'designed' or curated approaches to course sequencing, specifically examining how foundational knowledge can set the stage for higher-order understanding. We use a mixed-method approach, combining quantitative data analysis with qualitative insights from faculty and industry experts, to investigate the "Critical Paths of Business Education," and when the right time to introduce analytics could be.

4 - Analytically Astute: Training our Students to Understand Real World Phenomena

Srinivas Krishnamoorthy, Beedie School of Business, Simon Fraser University, Vancouver, BC, Canada

Students typically view analytics knowledge as a means to boost their career prospects rather than a way to understand real world phenomena. We can motivate our students to adopt a holistic analytics perspective by discussing examples from the happenings around us. We will explore such examples in the talk and offer tips for training our students to become analytically astute.

SB66

Regency - 604

Multitask and Federated Learning: Challenges of Optimization Under Heterogeneity

Invited Session

Artificial Intelligence

Chair: Hongyang Zhang, Northeastern University, Boston, MA, United States

Co-Chair: Yang Tianbao, Texas A&M University, College Station, TX, United States

1 - On the Effect of Defections in Federated Learning and How to Prevent Them

Han Shao, Toyota Technological Institute at Chicago, Chicago, IL, United States, Minbiao Han, Kumar Kshitij Patel, Lingxiao Wang

Federated learning is a machine learning protocol that enables a large population of agents to collaborate over multiple rounds to produce a single consensus model. There are several federated learning applications where agents may choose to defect permanently--essentially withdrawing from the collaboration--if they are content with their instantaneous model in that round. This work demonstrates the detrimental impact of such defections on the final model's robustness and ability to generalize. We also show that current federated optimization algorithms fail to disincentivize these harmful defections. We introduce a novel optimization algorithm with theoretical guarantees to prevent defections while ensuring asymptotic convergence to an effective solution for all participating agents. We also provide numerical experiments to corroborate our findings and demonstrate the effectiveness of our algorithm.

2 - Boosting Multitask Learning Through Higher-Order Task Affinities

Hongyang Zhang, Northeastern University, Boston, MA, United States

Multitask learning is a widely used paradigm for training models on diverse tasks, with applications ranging from graph neural networks to language model fine-tuning. Since tasks may interfere with each other, a key notion for modeling their relationships is {\em task affinity}. This includes pairwise task affinity, computed among pairs of tasks, and higher-order affinity, computed among subsets of tasks. Naively computing either of them requires repeatedly training on data pooled from various task combinations, which is computationally intensive. We present a new algorithm that can estimate task affinities without this repeated training.

The key idea is to train a ``base'' model for all tasks and then use a linearization technique to estimate the loss of any other model with a specific task combination. The linearization works by computing a gradient-based first-order approximation of the loss, using low-dimensional projections of gradients as features in a logistic regression trained to predict labels for the specific task combination. We show theoretically that the linearized model can provably approximate the loss when the gradient-based approximation is accurate, and empirically verify that on several large models. Then, given the estimated task affinity matrix, we design a semi-definite program for clustering to group similar tasks that maximize the average density of clusters.

We evaluate the performance across seven datasets, including multi-label classification on graphs, and instruction fine-tuning of language models. We show that our task affinity estimates are within 2.7% distance of the true affinities while needing only 3% of FLOPs compared to full training.

3 - A Multi-Objective Neural Architecture Search Framework for Optimizing Spiking Neural Networks Using NSGA-II

Esmat Ghasemi Saghand, university of south florida, Tampa, FL, United States, Susana Lai-Yuen

Spiking Neural Networks (SNNs) are emerging as an energy-efficient and biologically plausible alternative to traditional Artificial Neural Networks (ANNs). Unlike ANNs, SNNs process asynchronous and sparse spike information, reducing the need for expensive computations and lowering energy consumption. However, finding suitable SNN architectures is challenging and time-consuming due to their non-differentiable spiking dynamics. While Neural Architecture Search (NAS) techniques have been developed for ANNs, only a few studies have applied NAS to SNNs. In this study, we introduce MONAS-ESNN, a multi-objective NAS framework based on the NSGA-II algorithm, to automate the design of SNN architectures. MONAS-ESNN identifies the best architectures in terms of both accuracy and the number of generated spikes. During the search, we evaluate micro-level neural blocks that include relevant components for SNNs, such as connectivity patterns and feed-forward and backward connections. We introduce a new scoring function within MONAS-ESNN to predict the performance of candidate architectures without training, reducing the search time. The performance prediction is based on the diversity of spike activation patterns across different data samples and considers the binary activation pattern of parametric leaky integrate-and-fire (PLIF) neurons. MONAS-ESNN is evaluated across three image classification benchmarks. Our results demonstrate the effectiveness of the proposed framework in identifying SNN architectures that achieve higher accuracy with a lower number of spikes compared to existing methods. This work provides insights for improving the discovery of more effective and energy-efficient SNN architectures.

SB67

Regency - 605

Daniel H. Wagner Competition I

Award Session

Daniel H Wagner

Chair: James Cochran, The University of Alabama, Tuscaloosa, AL, United States

1 - JD.com Improves Fulfillment Efficiency with Data-driven Integrated Assortment Planning and Inventory Allocation

Shuo Sun, University of California,Berkeley, Berkeley, CA, United States, Zuo-Jun Shen, Yongzhi Qi, Hao Hu, ningxuan kang, Jianshen zhang, Xin Wang, Xiaoming lin

JD.com, a leading e-commerce company in China, has significantly improved its fulfillment efficiency to enhance customer satisfaction. JD.com uses a two-level inventory network with Regional Distribution Centers (RDCs) and Front Distribution Centers (FDCs). JD.com develops a decision support system that integrates optimization techniques and machine learning algorithms to decide the assortment at FDCs and daily inventory allocation from RDCs to FDCs. Implemented across JD.com’s network in China, it significantly enhances order fulfillment efficiency and reduces inventory and transfer costs.

2 - Optimizing Music Station Playlists on Broadcast Radio

J. Antonio Carbajal, iHeartMedia, Atlanta, GA, United States, Juan Ma, Nannan Chen, Mario Aboytes-Ojeda

iHeartMedia has pioneered the use of mathematical optimization in music playlist creation, developing an engine that generates 24/7 playlists for radio stations, subject to strategic scheduling goals and key business rules. Utilizing song meta data, the latest research and predictive analytics around song performance, and radio listenership data, the engine produces music playlists simultaneously optimizing strength and diversity. The engine has been successfully deployed to production and has proven effective in boosting the operational efficiency and quality of music scheduling processes.

SB68

Regency - 606

Undergradutate Operations Research Prize I

Award Session

Undergraduate Operations Research Prize

Chair: Rachel Cummings, Columbia University, New York, NY, United States

1 - Online Newsvendor Problem with Blackbox Predictions

Rohan Kulkarni, Columbia University, New York, NY, United States

We explore the application of algorithms with blackbox predictions to the online newsvendor problem, an established problem in inventory management. We introduce a no-regret algorithm, TEST, which leverages predictions to achieve exponential improvement in best-case regret bound compared to existing algorithms that do not use predictions. We also formulate novel regret-based consistency and robustness metrics to evaluate algorithm competitiveness in the online setting. TEST achieves a best-case consistency regret bound of O(T^(1/2)e^-T) while remaining sublinear in regret in the worst-case.

2 - Practical Bayesian Algorithm Execution via Posterior Sampling

Chu Xin (Cloris) Cheng, California Institute of Technology, Pasadena, CA, United States

We consider Bayesian algorithm execution, where the goal is to select points for evaluating an expensive function to best infer a property of interest. Observing that the property of interest for many tasks is a target set of points defined by the function, we derive PS-BAX, a simple and effective posterior sampling algorithm. Experiments show PS-BAX achieves competitive performance against baselines while being much faster to compute. Additionally, we show that PS-BAX is consistent under mild conditions. Our work thus provides new insights into posterior sampling and its application scope.

3 - Statistical Properties of Robust Satisficing

Zhiyi Li, Peking University, Beijing, China, People's Republic of

Robust Satisficing is an emerging robust optimization method. However, it lacks the statistical guarantees of its widely-studied counterpart, DRO. This paper addresses this gap by deriving two-sided confidence intervals for optimal loss and finite-sample generalization error bounds under Wasserstein distance, extending also to distribution shifts. For f-divergence, we establish an asymptotic upper bound on generalization error. Numerical experiments show that RS outperforms ERM in small-sample regimes and under distribution shifts, and shows less sensitivity to hyperparameter tuning than DRO.

4 - On Partial Optimal Transport: Revising the Infeasibility of Sinkhorn and Efficient Gradient Methods

Anh Duc Nguyen, National University of Singapore, Singapore, Singapore

We study the Partial Optimal Transport (POT) problem. We first investigate the infeasibility of the SOTA Sinkhorn algorithm for POT, degradeing its performance in applications. To this end, we propose a novel rounding algorithm for POT, and then provide a feasible Sinkhorn procedure. Our rounding algorithm also permits the development of two novel POT methods: Adaptive Primal-Dual Accelerated Gradient Descent and Dual Extrapolation which has the best complexity in the literature. We show the practicality of our algorithms on applications such as point-cloud registration and domain adaptation.

SB69

Regency - 607

Generative AI, Algorithmic Bias, and Dynamics in Digital Disruptions

Invited Session

eBusiness

Chair: Yi Gao, Texas Tech University, Lubbock, TX, United States

1 - Generative AI and Human Knowledge Sharing: Evidence from a Natural Experiment

Qili Wang, University of Florida, Gainesville, FL, United States, Yi Su, Kaiyu Zhang, Liangfei Qiu

Generative AI, known for its content creation capabilities, possesses great potential to reshape knowledge-sharing activities. By leveraging a unique policy of a leading online Q&A platform that introduces generative AI answers, we explore the effects of generative AI answers on human knowledge contribution efforts in knowledge-sharing platforms. Our findings demonstrate the positive effects of the generative AI answers on human knowledge contribution efforts, both in terms of quantity and length of the subsequent answers to a question. Notably, a more nuanced investigation into human-AI answer similarity indicates that generative AI answers prompt users to provide answers more aligned with the AI-generated answers, underscoring the AI-conformity effect rather than the AI-differentiation effect as the underlying mechanism. Moreover, we observe an increase in human experts' contribution following the introduction of generative AI answers, suggesting generative AI answers encourage rather than crowd out human experts' contribution to the platform. In a follow-up randomized experiment, we offer corroborative evidence for the results of empirical analysis. Furthermore, it provides preliminary evidence that labeling generative AI as the knowledge provider, rather than the content of generative AI answers, primarily drives the observed effects of introducing generative AI answers. Our research highlights the potential of generative AI to motivate human participation in knowledge generation and dissemination, adding to the burgeoning body of work on how generative AI influences human knowledge sharing.

2 - Wrong Model OR Wrong Practices? Mis-Specified Demand Model and Algorithmic Bias in Personalized Pricing

Qiaochu Wang, Carnegie Mellon University, Pittsburgh, PA, United States, Yan Huang, Param Singh, Kannan Srinivasan

The societal significance of fair machine learning (ML) cannot be overstated, yet quantifying algorithmic bias and ensuring fair ML remains a challenging task. One popular fair ML objective, equality of opportunity, requires equal treatment for individuals who are equally deserving, regardless of their group affiliation. However, determining who should be considered ``equally deserving" is a complex and critical aspect that directly affects the estimation of algorithmic bias. This paper emphasizes the importance of accurately measuring equal deservingness in order to accurately estimate algorithmic bias. To illustrate this, the paper examines the case of personalized pricing and shows that assuming a mis-specified model for equal deservingness can result in incorrect bias estimates. Using a detailed consumer data set from a large e-commerce platform, the paper demonstrates that when the correct consumer demand model is a non-sequential search model where consumers differ in their search costs based on gender, assuming a standard choice demand model or a traditional ML (e.g. Support Vector Machine) can lead to incorrect bias estimates. This research highlights the critical role that a proper model specification plays in achieving fair ML practices.

3 - Decoding Signal Discrepancies: Uncertainty and Exploration in Digital Music Consumption

Xinyu Fu, Georgia State University, Atlanta, GA, United States, Narayan Ramasubbu, Shadi Janansefat

Digital platforms deploy various signals to direct and shape user behaviors. Challenging the prevailing view that signal designs are crafted to reduce uncertainty, we leverage information gap theory to hypothesize that exposure to misaligned signals could, paradoxically, amplify content discovery behaviors. In particular, we focus on inconsistencies across signals from different sources, such as the disparity between users' social network preferences and overall platform trends. In a field experiment on a music streaming platform, participants were randomly assigned to one of four design scenarios that varied by the signal source and uncertainty level. Results indicate that users who received misaligned signals not only listened to more new music content, but also viewed a greater number of other users' profiles, relative to the other designs. We discuss the implications of information foraging behaviors of users for digital platforms and for further research on content discovery and user engagement.

4 - Balancing Biased Prices in Art: The Impact of Decentralized Nft Marketplace on Minority Artists

Agnes Yang, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, YoungJin Kwon

Web3 has ushered in a decentralized online ecosystem underpinned by blockchain technology, consequently reducing the influence of traditional gatekeepers and intermediaries. This development prompts our research inquiry: \textit{How does this emergent environment impact prevailing gender and racial biases?} To address this question, we collect artist-, NFT-, and transaction-level data from Superrare, a leading NFT art market, and revisit the concept introduced by Goldin and Rouse (2000), demonstrating how "blinding" musicians' identities affects audition results for female musician. We present a nuanced perspective on NFT markets: (i) White males continue to dominate the NFT art market in terms of supply volume; (ii) Female, Black, and Hispanic NFT artists experience discounts compared to their White and male counterparts; (iii) NFT artists who conceal their identity do not experience gendered and racialized valuations, supporting the notion of blind audition effect; (iv) Female and Black NFT artists who disclose their identity derive greater benefits from providing textual self-description compared to their White and male counterparts. The paper discusses theoretical contributions and managerial implications.

5 - Decoding Coding Advice: Generative AI is Thinking Outside the Box in Problem Solving

ZIYI ZHAO, Temple University, Philadelphia, PA, United States, Kanghyun Cho, Aleksi Aaltonen, Detmar Straub

Generative AI (GenAI) surpasses traditional search technologies by generating novel solutions that do not exist in the search space. As such, traditional understandings of problem-solving, attitudes toward algorithmic solutions, and related mechanisms do not necessarily apply in the context of GenAI. Using the Elaboration Likelihood Model (ELM) and Cognitive Load Theory (CLT), we investigate how programmers search, evaluate, and adopt GenAI- versus human-generated programming solutions in a sequential mixed-methods study. The first part of the study based on behavioral modeling reveals contradictory results whereby GenAI-generated solutions are perceived as higher quality, despite human sources being considered more credible. The apparent paradox is resolved in the second part of the study based on qualitative analysis of problem-solving sessions showing that improved quality can be attributed to reduced cognitive load when using GenAI. GenAI supports various problem-solving approaches, such as backward and iterative search, which foster human-in-the-loop to efficiently prompt GenAI for high-quality solutions. However, the cognitive ease of GenAI raises concerns about users uncritically accepting potentially flawed solutions, and that an effective human-in-the-loop approach necessitates appropriate prompt representation, by which we refer to the iterative structuring of queries to guide GenAI toward more accurate and relevant solutions. This paper contributes to the understanding of human-GenAI collaboration in problem-solving.

SB70

Regency - 701

Data-driven Analytics for Applications in Water and Food Supply

Invited Session

Energy: Natural Resources

Chair: Victoria Chen, The University of Texas at Arlington, Arlington, TX, United States

1 - A Design and Analysis of Computer Experiments Approach to Water Distribution Network Rehabilitation Optimization

Uthman Kareem, University of Texas at Arlington, Arlington, TX, United States, Jay Rosenberger, Victoria Chen, Mohsen Shahandashti, Raghavendra Kumar Punugu

Determining seismic vulnerabilities of water network modeling is critically important to ensure access to water in the face of an earthquake. In this research, we apply design of experiments from a number theoretic method (NTM) and a seismic modeling approach to create several post-disaster scenarios of a network. The variables in the design are the pipes while their levels are different states such as leaks, breaks or unaffected after network disruption. The same sample points from NTM are used to generate different rehabilitation policies through a ranking algorithm. These rehabilitation policies are applied on each scenarios and evaluated using computer simulation. The simulation results present opportunity for analysis as a way of generating better alternative improvement policies.

2 - A Cvar-Var-Standard Deviation Multi-Objective Risk Optimization for Enhancing Seismic Resilience in Water Pipe Network Rehabilitation

Raghavendra Kumar Punugu, The University of Texas at Arlington, Arlington, TX, United States, Jay Rosenberger, Victoria Chen, Uthman Kareem, Mohsen Shahandashti

Risk measures such as value-at-risk (VaR), conditional value-at-risk (CVaR), and standard deviation are among the popular risk measures widely used in financial risk management with applications in portfolio management, risk capital management, performance attributions and many more. Although these are predominantly used in quantitative finance, in this talk they are employed for enhancing water pipe network resilience to seismic events, which involves high uncertainty. VaR and CVaR at a given confidence level for a random variable X, representing the network's System Serviceability Index (SSI), are defined as the lower α-tail percentile and the conditional expectation below VaR, respectively. In this study, we propose using VaR, CVaR, and standard deviation as performance metrics to evaluate rehabilitation policies as a function of SSI for decision-making under seismic uncertainty. Using VaR-CVaR-Standard Deviation, we develop a multi-objective risk optimization model to increase the value realization from VaR-CVaR and minimize the value realization from standard deviation, generating a set of nondominated policies

3 - Machine Learning for Predicting Crop Yield Based on Soil Properties

Darya Abbasi, University of Texas at Arlington, Arlington, TX, United States, Amanda Ashworth, Phillip Owens, Edwin Winzeler, Kharel Tulsi, Yuan Zhou

Soil, as the primary agricultural medium, holds a crucial position in farming practices, and understanding the intricate relationship between soil properties and crop yield stability is vital for effective agricultural management. This study aims to use machine learning (ML) to predict crop yield stability based on comprehensive soil data analyses. The research was carried out across four sites in Ottawa County, Oklahoma, USA, encompassing a comprehensive dataset featuring yield data for corn, soybean, canola, and wheat, alongside soil nutrient properties across varying depths.
This abstract covers two main topics. First, we define the measurement of yield stability and discuss the process for merging multi-source data at a high spatial resolution within precision agriculture. Then, we discuss the ML techniques used to predict crop yield stability and explore their correlations with soil properties. The findings highlight the most effective approaches for linking soil characteristics to crop yield stability, emphasizing the importance of consistent and reliable crop production for long-term agricultural sustainability and food security.

4 - Designing a Crowd Logistics Network of a Regional Food System by Analyzing An Agent-Based Model Using Sequential Sampling

Victoria Chen, The University of Texas at Arlington, Arlington, TX, United States, Preetam Kulkarni, Caroline Krejci

Regional food systems involve farmers selling locally produced food to customers and it would benefit farmers if they can expand their market reach. It also gives customers access to fresh produce and promotes better health. Small scale farmers tend to be involved in transporting their produce to the destination as they can not afford hiring a third party logistics service as it is expensive. Moreover, farmers are also concerned about the safety of the produce while it is being transported. Crowd logistics is a potential solution to this problem as it could help farmers by providing an affordable and safe means of transportation given that the participants on the crowd logistics network are vetted and credible. This research aims to develop methods to evaluate a crowd logistics network using agent based modeling as these models tend to be computationally intensive. We propose the use of sequential sampling as a means to reduce the number of simulation runs and still be able to evaluate the factors influencing the performance of a crowd logistics network as it's success is important to ensure continued service to farmers.

SB71

Regency - 702

Online Learning and Data-driven Decision Making: Recent Advances and Applications

Invited Session

Data Mining

Chair: Tanapol Kosolwattana, University of Houston, Houston, TX, United States

Co-Chair: Ying Lin, University of Houston, Houston, TX

1 - A Tensor-Based Approach for Monitoring Partially Observed High-Dimensional Data Stream

Junghee Pyeon, Georgia Institute of Technology, Atlanta, GA, United States, Kamran Paynabar

Many types of high-dimensional (HD) data stream have become available to engineers in a number of fields thanks to the development of measuring equipment as well as communication technology. In practice, however, a variety of reasons such as insufficient transmission bandwidth, time constraints, and surging costs, forces practitioners to deal with partially observed information in order to examine whether their processes are currently working without any anomalies. To address these challenges, we propose a novel framework capable of effectively monitoring incomplete streaming HD data by capturing the inherent auto-/cross-correlation. To be specific, our approach involves two key components: (i) delay-embedding transformation along with the time axis, which has been widely used for missing value imputation, to generate the high-dimensional but low-rank tensor and (ii) Tucker decomposition to fragment the tensor into invariant factorizing matrices and time-dependent core tensors, which are then monitored by statistical process control methodologies. Through intensive simulation experiments and a case study, we evaluate the performance of our procedure and compare it with existing methods.

2 - Fcom: a Federated Collaborative Online Monitoring Framework via Representation Learning

Ying Lin, University of Houston, Houston, TX, United States, Tanapol Kosolwattana, Huazheng Wang, Raed Kontar

Monitoring a large population of dynamic processes with limited monitoring resources poses a significant challenge across various industrial sectors due to 1) the inherent disparity between the limited monitoring resources available and the large population of processes to be monitored and 2) the unpredictable and heterogeneous dynamics inherent in the progression of these processes. Online learning approaches, commonly referred to as bandit methods, have demonstrated notable potential to solve this problem by dynamically allocating limited resources, effectively balancing the exploitation of processes yielding high rewards, and the exploration of uncertain processes. However, most online learning algorithms are designed under 1) a centralized setting that requires data sharing across processes to obtain an accurate prediction or 2) a homogeneity assumption that estimates a single global model from the decentralized data. To overcome these limitations and facilitate online learning in a heterogeneous population under a decentralized setting, we propose a federated collaborative online monitoring method. Our approach captures the latent representative models inherent in the population through representation learning and designs a novel federated collaborative UCB algorithm to estimate the representative models from sequentially observed decentralized data. This allows an informed strategy for monitoring resource allocation. The efficiency of our method is illustrated through theoretical analysis, simulation studies, and decentralized cognitive degradation monitoring in Alzheimer's disease.

3 - MT-RAM: Multi Task-Recurrent Attention Model for Partially Observable Image Anomaly Classification and Localization

Jie Guo, Tsinghua University, Beijing, China, People's Republic of, Chen Zhang

With the rapid development of the digital manufacturing industry, the nature of quality data has transformed from simple univariate or multivariate characteristics to big data comprising multimedia elements such as images and videos. The utilization of image data for automated monitoring and anomaly detection has gained significant attention in recent years, which also poses new and complex challenges. A critical challenge is the substantial demand for sensing and computation resources. When these resources are limited, only a fraction of the image data can be observed and analyzed. Hence, adaptive sampling becomes imperative to select the most informative pixels that effectively capture anomaly information. In this paper, we propose a novel recurrent neural network framework named Multi Task-Recurrent Attention Model (MT-RAM) which incorporates adaptive sampling for anomaly classification and localization in partially observable image data. MT-RAM emulates human-like perception by generating a sequence of glimpses to comprehend the image, with the location of each glimpse depending on the information gleaned from previous glimpses. Thorough numerical studies and case studies are conducted to evaluate the performance of MT-RAM in comparison to state-of-the-art adaptive-sampling-based anomaly detection methods.

4 - Fourier Learning with Cyclical Data

Zhihan Xiong, University of Washington, Seattle, WA, United States, Yingxiang Yang, Tianyi Liu, Taiqing Wang, Chong Wang

Many machine learning models for online applications, such as recommender systems, are often trained on data with cyclical properties. These data sequentially arrive from a time-varying distribution that is periodic in time. Existing algorithms either use streaming learning to track a time-varying set of optimal model parameters, yielding a dynamic regret that scales linearly in time; or partition the data of each cycle into multiple segments and train a separate model for each---a pluralistic approach that is computationally and storage-wise expensive. In this paper, we have designed a novel approach to overcome the aforementioned shortcomings. Our method, named ``Fourier learning", encodes the periodicity into the model representation using a partial Fourier sequence, and trains the coefficient functions modeled by neural networks. Particularly, we design a Fourier multi-layer perceptron (F-MLP) that can be trained on streaming data with stochastic gradient descent (streaming-SGD), and we derive its convergence guarantees. We demonstrate Fourier learning's better performance with extensive experiments on synthetic and public datasets, as well as on a large-scale recommender system that is updated in real-time, and trained with tens of millions of samples per day.

5 - Bandit Optimisation under Fused Gaussian Processes

Muyun Lu, University of Houston, Houston, TX, United States, Huazheng Wang, Ying Lin

In many situations across computational science and engineering, there are tasks to maximize an expensive to estimate black box function of a system (target function). To save the cost, in many cases, cheap approximations to the target function can be available to gain partial information which helps us identify the optimum at the target function, which are known as multi-fidelity bandit optimisation. However, existing multi-fidelity bandit optimisation models ignore the latent structure among fidelities in reality, where the observations in target fidelity and its approximations can be correlated by a fusion model. To solve this problem, we develop a novel Fused Gaussian Process-based Upper Confidence Bound (FGP-UCB) method which leverages the low fidelities to gain partial information for uncertainty reduction through the fusion model and facilitates the target fidelity to identify the optimum quickly. Comparing to the state-of-art multi-fidelity GP-UCB approach, our model does not constrain the functions in different fidelities via a uniform bound, which is more flexible to fit the real data. Effectiveness of the proposed method is demonstrated via both theoretical analysis and empirical studies.

SB72

Regency - 703

Machine Learning Aided Causal Inference

Invited Session

Data Mining

Chair: Srikar Katta, Duke University, Durham, NC, United States

1 - Assessing Robustness to Unobserved Confounding by Combining Experimental and Observational Data

Quinn Lanners, Duke University, Durham, NC, United States

Combining data from a randomized experiment with observational study data can offer considerable improvements in efficiency and generalizability; however, this comes at a great cost, as it typically sacrifices the robustness to confounding which is the hallmark of randomization. To avoid biased estimates when using observational data, we rely on a different assumption, referred to as study exchangeability. While this assumption may be more valid in certain scenarios, it is a difficult assumption to validate and violations in study exchangeability can have equally as detrimental effects as unobserved confounding. In this talk, we discuss how we handle the assumptions of conditional ignorability and study exchangeability more carefully. Particularly, we present a partial identification approach to heterogenous treatment effect estimation when combining randomized experiment and observational study data under potential violations to both conditional confounding and study exchangeability.

2 - Bayesian Causal Synthesis for Meta-Inference on Heterogeneous Treatment Effects

Kenichiro McAlinn, Temple University, Philadelphia, PA, United States, Shonosuke Sugasawa, Kosaku Takanashi, Edoardo Airoldi

The estimation of heterogeneous treatment effects in the potential outcome setting is biased when there exists model misspecification or unobserved confounding. As these biases are unobservable, what model to use when remains a critical open question. In this paper, we propose a novel Bayesian methodology to mitigate misspecification and improve estimation via a synthesis of multiple causal estimates, which we call Bayesian causal synthesis. Our development is built upon identifying a synthesis function that correctly specifies the heterogeneous treatment effect under no unobserved confounding, and achieves the irreducible bias under unobserved confounding. We show that our proposed method results in consistent estimates of the heterogeneous treatment effect; either with no bias or with irreducible bias. We provide a computational algorithm for fast posterior sampling. Several benchmark simulations and an empirical study highlight the efficacy of the proposed approach compared to existing methodologies, providing improved point and density estimation of the heterogeneous treatment effect, even under unobserved confounding.

3 - Auditing Fairness under Unobserved Confounding

Michael Oberst, Johns Hopkins University, Baltimore, MD, United States, Yewon Byun, Dylan Sam, Zachary Lipton, Bryan Wilder

Inequity in resource allocation has been well-documented in many domains, such as healthcare. Causal measures of equity / fairness seek to isolate biases in allocation that are not explained by other factors, such as underlying need. However, these fairness measures require the (strong) assumption that we observe all relevant indicators of need, an assumption that rarely holds in practice. For instance, if resources are allocated based on indicators of need that are not recorded in our data ("unobserved confounders"), we may understate (or overstate) the amount of inequity.

In this talk, I will present work demonstrating that we can still give informative bounds on certain causal measures of fairness, even while relaxing (or even eliminating) the assumption that all relevant indicators of need are observed. We use the fact that in many real-world settings (e.g., the release of a new treatment) we have data from prior to any allocation, which can be used to derive unbiased estimates of need. This result is of immediate practical interest: we can audit unfair outcomes of existing decision-making systems in a principled manner. For instance, in a real-world study of Paxlovid allocation, we show that observed racial inequity cannot be explained by unobserved confounders of the same strength as important observed covariates.

4 - Covariate-Assisted Inference on Partially Identified Causal Effects

Asher Spector, Stanford University, Stanford, CA, United States, Wenlong Ji, Lihua Lei

Many causal estimands are only partially identifiable since they depend on the unobservable joint distribution between potential outcomes. Stratification on pretreatment covariates can yield sharper partial identification bounds; however, unless the covariates are discrete with relatively small support, this approach typically requires consistent estimation of the conditional distributions of the potential outcomes given the covariates. Thus, existing approaches may fail under model misspecification or if consistency assumptions are violated. In this study, we propose a unified and model-agnostic inferential approach for a wide class of partially identified estimands, based on duality theory for optimal transport problems. In randomized experiments, our approach can wrap around any estimates of the conditional distributions and provide uniformly valid inference, even if the initial estimates are arbitrarily inaccurate. Also, our approach is doubly robust in observational studies. Notably, this property allows analysts to use the multiplier bootstrap to select covariates and models without sacrificing validity even if the true model is not included. Furthermore, if the conditional distributions are estimated at semiparametric rates, our approach matches the performance of an oracle with perfect knowledge of the outcome model. Finally, we propose an efficient computational framework, enabling implementation on many practical problems in causal inference.

SB73

Regency - 704

Data Mining Best Paper Competition (Student Track)

Award Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

Co-Chair: Andi Wang, University of Wisconsin-Madison, Madison, WI, United States

1 - Distributionally and Adversarially Robust Logistic Regression via Intersecting Wasserstein Balls

2 - MOSS: Multi-Objective Optimization for Stable Rule Sets

3 - Statistical and Computational Guarantees of Kernel Max-Sliced Wasserstein Distances

4 - A Bayesian Jump Model-based Pathwise Sampling Approach for Online Anomaly Detection

Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States, Miao Bai, Di Wang, Xiaochen Xian

Moving vehicle-based sensors (MVSs) have received growing attention for real-time anomaly detection in various applications such as wildfire and oil spill detection. To tackle challenges due to the spatial covariance structure among observations, uncertainties under partial observations, as well as the physical MVS movements, we propose a Bayesian jump model-based pathwise sampling approach to detect abrupt changes in an area of interest in real time using MVSs. Specifically, we integrate a jump-model based Bayesian scheme, the upper confidence bound algorithm, and mathematical optimization in a unified manner to exploit spatial correlation and real-time partial observations, to handle uncertainties of noisy observations, limited observability, and anomaly occurrences, and to adaptively coordinate the routes of multiple MVSs for quick anomaly detection. We perform theoretical investigations and conduct simulations to confirm the exceptional effectiveness of the method we propose. A case study for early wildfire detection demonstrates that our proposed method outperforms benchmark methods, which contributes to the reduction of the area of affected land and wildfire-related costs.

SB74

Regency - 705

Issues in Energy Market Modeling

Invited Session

ENRE: Electricity

Chair: Joseph Duggan, University of Dayton, Dayton, OH, United States

Co-Chair: Ramteen Sioshansi, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Stochastic Capacity Expansion Modeling Considering Interconnection Queue Risk

Yiwen Wang, University of Massachussets, Amherst, Amherst, MA, United States

Interconnection queue is a list of projects awaiting impact studies before getting approved for grid connections. Long delays and high withdrawal rates have been observed in the interconnection queues in many regions of the US. The delays and withdraws predominantly affect renewable and battery-hybrid projects, potentially impeding the energy transition. Few studies have examined how risks associated with interconnection queues influence investment decisions of the energy system. This study employs a multi-year stochastic capacity expansion model that accounts for the uncertainty in lead times and withdrawal rates of new projects. We utilize the Expected Value of Perfect Information (EVPI) and the Value of Stochastic Solution (VSS) to assess the impacts of the interconnection queue congestion on deep decarbonization efforts. The results aim to inform how interconnection policy could affect deep decarbonization.

2 - Uniqueness of Nash-Cournot Equilibria in an Oligopolistic Energy Market with Demand Uncertainty and Uncertain Renewable Generation

Benjamin Chaiken, Parametric Portfolio Associates, Seattle, WA, United States, Joseph Duggan

This work explores the uniqueness of Nash-Cournot equilibria in wholesale electricity markets with both thermal and renewable generation in the presence of both uncertain demand uncertainty and uncertain renewable generation. In a previous work, the authors expand upon the uniqueness results proven by Lagerl ̈of (2006) to incorporate the presence of renewable generation into the market in the manner of Acemoglu et al. (2017). The model of Acemoglu et al. (2017) incorporates a constant amount of renewable generation, with zero marginal cost, distributed among all firms. The present work extends this model to include uncertainty over the amount of renewable generation available in any given period, and extends the prior uniqueness results to study the effects of uncertain renewable supply on the multiplicity of equilibiria.

3 - Designing Carbon Policy with Profit-Maximising Energy Storage

Afzal Siddiqui, Stockholm University, Kista, Sweden, Ramteen Sioshansi

In a storage-enabled renewable-thermal power system, carbon emissions still require mitigation based on their social cost of damage. A central-planning solution internalises the environmental externality by curbing thermal production in proportion to the marginal cost of damage. Consequently, in a two-period setup, the price in the peak (second) period increases by more than that in the off-peak (first) period. In particular, the rates at which the prices increase differ by the inverse of the storage device’s efficiency. Thus, an infinitesimal increase in the slope of the marginal cost of damage curbs thermal generation in both periods without affecting storage-enabled renewable operations. Moreover, a degradation in storage efficiency has a monotonic impact on the marginal cost of damage. By contrast, in a decentralised electricity industry, a bi-level approach is necessary in which a welfare-maximising policymaker at the upper level sets a carbon tax to impose on industry at the lower level. For a completely efficient storage device, the carbon tax in this bi-level setting renders the first-best outcome. However, in the presence of an inefficient storage device, an infinitesimal increase in the carbon tax induces prices in both periods to increase at the same rate. As a result, the storage-enabled renewable operator shifts energy to the off-peak period in order to offset the loss in the value of stored energy. Hence, relative to a centralised industry’s marginal cost of damage from emissions, the carbon tax in a decentralised industry is lower and may be nonmonotonic in the storage device’s inefficiency.

4 - The implications of storage degeneracy on power system optimization models

Thomas Klatzer, Graz University of Technology, Graz, Austria, Sonja Wogrin, Ramteen Sioshansi

With the ongoing transition towards decarbonized power systems, storage technologies are becoming increasingly important to mitigate the variability of renewable generation, i.e. solar and wind, and to achieve modest total system cost. In the context of power system optimization models (PSOMs), storage technologies can contribute to degeneracy in the decision variables, i.e. when a storage unit is charged and its corresponding state of charge, which implies that the determined values of the decision variables are not unique. Against this background, we formulate a PSOM with a highly flexible temporal structure to study degeneracy of both short-term and long-term storage technologies using a full chronological representation of time and representative periods. As a unique model feature, the representative periods framework enables modeling of long-term storage behavior without sacrificing the computational benefits of such a framework. In our talk, we will discuss the impact of degeneracy of short-term and long-term storage technologies on decision variables, computational performance and local marginal prices and showcase approaches to reduce storage degeneracy.

5 - Routing India's thermal generation towards Net Zero with CCS: optimal planning and development policy implications

Olivier Massol, CentraleSupélec & City, University of London, Gif sur Yvette, France, Adrien Nicolle, Stephanie Monjon

India’s CO2 emissions are among the highest globally, largely due to its extensive coal-fired power plant fleet. Despite the commitment to achieving net-zero emissions, India has not planned a coal phase-out and continues to expand its coal power capacity to meet the nation's rapidly growing consumption of electricity. Carbon Capture and Storage (CCS) technologies offer a viable solution to mitigate these emissions while preserving the country's ambition to monetize its domestic coal resources. However, CCS deployment in India is still embryonic as no Indian coal-fired power plant currently employs CCS. The puprose of this study is to: (i) examine the optimal deployment of CCS in India's power sector, (ii) analyze its economics to evaluate the associated investment gap, and (iii) evaluate policy options aimed at accelerating CCS adoption. We use a network optimization approach to design a cost-effective pipeline network for transporting CO2 from coal power plants to storage sites across India. Our analysis shows that a significant portion of the transport infrastructure is consistent across various scenarios, highlighting strategic priorities for infrastructure investments. This insight is crucial as Indian authorities and development agencies consider subsidizing CCS infrastructure, providing clear guidance on which pipelines should receive priority investment. Our research offers valuable insights into the economic and policy frameworks necessary for the efficient and accelerated deployment of CCS technology, supporting India's transition towards a sustainable energy future.

SB75

Regency - 706

Planning the Electric Grid for Climate Change

Invited Session

ENRE: Energy-Climate

Chair: Sofia Taylor, University of Wisconsin-Madison, Madison

Co-Chair: Line Roald, University of Wisconsin - Madison, Madison, WI, United States

1 - Winter vulnerabilities to weather and climate in future power systems

Ana Dyreson, Michigan Technological University, Houghton, MI, United States, Shelbie Wickett

Changes in weather and climate impact electric power system vulnerabilities in myriad ways that depend also depend on technology changes in supply and demand. This talk focuses on planning electric power systems for winter weather. In northern climates, where winter heating loads are significant, adoption of electric heat pumps will change the seasonal electricity demand at the same time that solar photovoltaics (PV) become a critical part of resource supply in those regions. We use high resolution daily snow depth data to estimate how snow will affect the availability of future solar PV resource supply. Using metrics including capacity credit, we overlay these supply changes with winter and spring peak demand for electricity for heating to determine how cold and snowy weather will overlay to affect future resource adequacy. These analyses highlight key springtime periods for resource adequacy and reliability assessments and underscore the need to continue to understand how snow affects solar PV systems and how snow shedding from solar PV can be better mitigated and forecast for operational planning.

2 - Computational pipeline predicts solar and wind energy availability under various climate change projections

Matthew Signorotti, Lawrence Livermore National Laboratory, Livermore, CA, United States

Background/objectives

Renewable and low-carbon energy sources should power the grid to minimize future climate change. This project seeks to predict climate dependent generation availability, such as from solar and wind farms, under various future climate projections.

Approach/activities

I create a computational pipeline for working with large numeric objects such as scientific array data. Features include streamlining reading inconsistent netCDF files while performing various validation checks, lazy evaluation, eager garbage collection, and automatic unit conversion. The pipeline is used to run renewable power simulations with the System Advisor Model (SAM) from the National Renewable Energy Laboratory (NREL) for different netCDF climate datasets, from past to future data, with various gridding schemes, spatial and temporal resolutions, variables, units, climate projection scenarios, and data sources. The pipeline builds upon numerous climate simulations, which themselves include just climatological information. It generates future power simulation data that can inform sustainable energy policy investment decisions.

Results/lessons learned

I present empirical results on real climate datasets and implications for future renewable power availability.

3 - Optimizing Power Line Undergrounding Decisions under Wildfire Risk and Varying Weather Scenarios

Sofia Taylor, University of Wisconsin - Madison, Madison, WI, United States, Amelia Musselman, Line Roald, Jean-Paul Watson

The threat of wildfire ignitions from electric power equipment has led utilities to increasingly turn to preemptive power shutoffs, which, while effective in reducing grid-induced wildfire risk, can cause significant load loss. Undergrounding power lines is an alternative strategy for preventing grid-induced wildfires. However, undergrounding lines is costly, so an effective undergrounding plan must balance the cost of load loss with the cost of undergrounding lines. We propose a robust optimization model to identify power lines to underground to maximize load served while limiting wildfire risk across a range of wildfire risk and weather scenarios. Since solving this problem may be computationally heavy for large power systems and many operating scenarios, we present a delayed constraint generation algorithm to iteratively add scenarios until an optimal solution is found. We evaluate the performance of this robust optimization framework on the RTS-GMLC system with scenarios representing a year of operating conditions.

4 - Resilient Capacity Planning with EPRI’s Climate READi Initiative

Jesse Bukenberger, Electric Power Research Institute, Pittsburgh, PA, United States

Electric grids have always been exposed to extreme events with hazards such as extreme heat, cold, and storm-force winds, but the nature of these hazards is shifting with the climate. Planners need more rigorous ways to plan a resilient system. This work gives utilities a systematic approach to integrate resilience goals into established long-term capacity planning practices.

One path to resilience is through hardening. Individual assets can often be hardened with outfits that protect against some, but usually not all, hazard impacts. In the system more broadly, resilience can also be improved through more traditional capacity planning decisions that include changing the size, location, and timing of new generation and transmission investments. These investments can bring benefits beyond resilience, such as reduced operating costs and emissions. The plan that best meets all the planning requirements will likely include a balance of these options, and finding such a plan requires an integrated planning model.

The presented model optimizes investments in generation, transmission, and hardening options. Extreme events with many different hazards are considered alongside normal operating conditions. Specific resilience targets are included as planning criteria. The resulting solution is the least-cost expansion plan that meets the planner's specified resilience and other planning targets. We demonstrate the approach and highlight some resulting insights on a 4,894-node synthetic Texas power system with extreme heat, cold, and hurricane events.

SB76

Regency - 707

Multi-Objective Decision Support Systems for Coping with Deep Uncertainties

Invited Session

ENRE: Environment and Sustainability

Chair: Chao Wang, Beijing University of Technology, 100 PING LE YUAN, Beijing, 100124

1 - Resilience strategies in an intertwined supply network: Mitigating the vulnerability under disruption ripple effects

Chao Wang, Beijing University of Technology, Beijing, China, People's Republic of

Intertwined supply network (ISN) is a complex system consisting of multiple interconnected supply chains orchestrated by different focal firms. The inherent interconnectedness of an ISN makes it vulnerable to localized disruptions, which can propagate through intricate supplier-buyer relationships, causing unpredictable disruption ripple effects that exacerbate vulnerabilities. While existing literature provides valuable insights into the vulnerability of supply networks with a single focal firm, less attention has been given to ISN comprising multiple competing focal firms. This study proposes an agent-based model to simulate the diffusion of disruption ripple effects in a realistic ISN. The model is used to assess the relative effects of firms' structural and operational attributes on vulnerability and to compare the effectiveness of three mainstream resilience strategies—resistance, adaptation, and recovery—in mitigating the overall vulnerability of ISN. Furthermore, the interaction of resilience strategies initiated by different focal firms in the ISN is explored. The results suggest that focal firms should prioritize their own structural attributes and those of their neighboring firms to address vulnerabilities effectively. Moreover, focal firms can identify partners with higher reciprocity and engage in cooperation for resilience investment. This study contributes to the understanding of ISN vulnerability and provides practical insights for developing effective resilience strategies in complex supply networks.

2 - A Decision Support System for Improving Drought Resilience under Climate Change

Abdullah Konak, Penn State Berks, Reading, PA, United States, Michael Jacobson, Daniel Winstead

Environmental risk modeling involves using indicators to measure extreme climate events and their impact, as well as coping mechanisms to reduce their effects. Index-based approaches make quantifying, comparing, and monitoring risks associated with natural hazards easier because a large set of indicators can be summarized into a few key performance indicators. This work will present a decision-support system (DSS) developed from a Belmont international collaborative research project to help policymakers devise and evaluate drought response and adaptation strategies based on a composite index-based approach. As drought events are becoming more frequent and unpredictable due to climate change, there is a need for decision support systems that incorporate risk assessment tools to help policymakers make decisions under uncertainty with incomplete data. We have developed a decision-support system framework that can integrate diverse climatic, biophysical, technological, and socioeconomic indicators to evaluate the drought resilience of at-risk communities. The framework considers complex relationships among these various indicators, combined with expert opinions from these diverse fields, to provide a what-if analysis for holistic policy recommendations.

3 - Should we stay or should we leave? Multi-objective tradeoffs driven by choices between beach nourishment and managed retreat under deep uncertainty.

Prabhat Hegde, Dartmouth College, Hanover, NH, United States, Adam Pollack, Vikrant Vaze, Klaus Keller

In some low-lying coastal areas around the world, decision-makers are beginning to consider "managed retreat" of human populations to adapt to sea-level rise. One of the main challenges in designing a managed retreat strategy is determining when to trigger retreat. Decision-makers lack tools and scientific understanding to evaluate whether the benefits of interim response strategies, like beach nourishment, outweigh the costs of waiting longer to trigger retreat. The current academic state of the art identifies the optimal timing of retreat for maximizing net present value with an assumption of perfect knowledge about climate and economic parameters. However, when evaluating strategies that look at timing sequences of beach nourishment and managed retreat, this approach remains silent on certain key considerations: First, there are trade-offs between key planning objectives such as benefits, investment costs, damage costs, and reliability. Second, this approach does not address how deep uncertainties surrounding climate and economic parameters influence these trade-offs and decision-making. Third, from a practical standpoint, the approach overlooks the viability of a strategy being financed by ignoring government benefit-cost ratio tests. Here, we introduce and operationalize an open-source, many-objective robust decision-making quantitative framework that fills these gaps. We identify the planning horizon and uncertainty ensemble required to achieve numerical convergence on the timing of retreat. We demonstrate that accounting for multiple objectives and uncertainty helps us identify robust strategies that also pass the government cost-benefit ratio tests more often.

4 - Mapping crop residue cover by integrating satellite time-series imagery and machine learning

Leticia Santos, North Carolina State University, Raleigh, NC, United States, Daniela Jones, Brian Lamb, Dean Hively, Jyoti Jennewein, Alison Thieme, Chris Reberg-Horton

Conservation tillage is a pivotal sustainable agricultural practice that enhances soil health and biodiversity by leaving crop residue on the field. This residue acts as a protective layer, preventing erosion, conserving soil moisture, and suppressing weeds, while also aiding in soil carbon sequestration to mitigate climate change. However, assessing its adoption accurately on a large scale poses significant challenges due to the time-consuming and labor-intensive nature of traditional ground-level surveys, which also lack scalability. To address these challenges, this study utilizes aerial and space-borne sensors, including shortwave infrared (SWIR) satellite imagery combined with machine learning algorithms, to efficiently collect high-resolution data. Specifically, the research maps and quantifies crop residue in croplands across the Delmarva Peninsula. By scaling ground-truthed high-resolution (4m pixels, 144 km2 footprint) crop residue cover maps derived from WorldView-3 SWIR imagery to more widely available, medium resolution Landsat images (30m pixels, 33,300 km2 footprint), we apply advanced machine learning techniques such as random forest, gradient boosting, and neural networks. The study incorporates inventories of conservation tillage on agricultural lands using imagery from satellite sensors and a novel, low-cost, scale-neutral tractor-mounted camera system developed by the PSA team. This approach not only quantifies soil cover residue, which may include cover crops, but also empowers government agencies and agricultural stakeholders with high-throughput techniques to measure, monitor, report, and verify tillage intensity practices on large-scale agricultural lands. This research are expected to enhance decision-making processes for stakeholders, policymakers, and farmers, and incentivize climate-smart agricultural practices that reduce greenhouse gas emissions.

5 - The Development of Machine Learning Models for Assessing In-season Sweetpotato Root Growth and Crop Yield Estimates

Shana McDowell, North Carolina State University, Raleigh, NC, United States, Daniela Jones, Michael Kudenov, Shelly Hunt

Sweetpotatoes vary widely in shape and size, and consumers prefer particular characteristics over others. Optimizing the yield to match consumer preferences would increase growers' profits and reduce waste. We used machine learning algorithms to highlight the environmental factors, harvesting, planting, and application decisions that contribute to sweetpotato shape and size throughout the growing season. In this process, we also gained a better understanding of the sweetpotato developmental stages through root images of sweetpotatoes as input into computer vision algorithms. We collaborated with a large sweetpotato grower, packer, and distributor in North Carolina, managing nearly fifteen thousand acres. Data was collected from several on-farm locations scattered over 2,500 square miles across three growing seasons. Sweetpotato weight and length-to-width ratio were predicted using Linear Regression, Support Vector Machine, Random Forest, and XG Boost models. Data splitting included random split, split between fields, and split between seasons. Results show that important predictors of sweetpotato weight are days after planting, average relative humidity, and average temperature. Predictors for length-to-width ratio include days after planting, the time between planting and liquid fertilizer application, and maximum wind speed. XG Boost performed best for predicting weight (MSE = 0.24, R^2 = 0.99), and Random Forest performed best for predicting length-to-width ratio (MSE = 0.01, R^2 = 0.98). As part of our translational research, we created a dashboard displaying in-field data to support stakeholder decision-making. This analysis aids in developing farm-to-market decision models, optimizing crop yield and crop quality, and better utilizing growers' resources.

SB77

Regency - 708

Network Models in Optimization and Their Applications

Invited Session

Computing Society

Chair: Yajun Lu, Jacksonville State University, Jacksonville, AL, United States

1 - The Longest Path Problem in Perturbed Directed Acyclic Graph and Its Application in Manufacturing Scheduling

Golshan Madraki, Clarkson University, Potsdam, NY, United States, Seyedamirabbas Mousavian

We proposed *an algorithm* to accelerate the calculation of the length of the longest path in a directed acyclic graph where there are multiple added and deleted edges. Our algorithm updates the affected nodes instead of calculating the length of the longest path from scratch. Our algorithm has an application in Manufacturing system scheduling problems. It can be used to accelerate the scheduling heuristics (such as Genetic Algorithm, Simulated Annealing, etc.) in manufacturing systems) because these heuristics use a graph to model the system and calculate the makespan, i.e., longest path in the graph, during each iteration after scheduling perturbations. These algorithms iteratively perturb the graph and recalculate the makespan in each iteration until a satisfactory schedule is determined. These scheduling perturbations can be modeled by added and deleted edges in the corresponding graph of the manufacturing systems.

2 - Robust Supply Chain Network Design: Maximizing Matching Size with Adversarial Node Deletion

Eugene Ang, National University of Singapore, Singapore, Singapore, Yifan Feng

Supply chain disruptions, such as natural disasters and geopolitical events, pose significant challenges to businesses and economies worldwide. In this paper, we extend the existing work from Feng et al., 2023 by considering adversarial node deletion and investigate how appropriate graph design can enhance the robustness of supply chain networks, by maximizing its matching size in the residual graph upon adversarial node deletion. In brevity, this is a (n,m,k) problem, where n and m are the order and the size of the designed graph G, respectively, and k is the number of adversarially deleted nodes. When n and k are of the same parity, we show that the optimal graphs G^* have certain graph properties such as connectedness and minimum degree of k + 1. When n and k are of different parity, we obtain bounds on the size of G^* by analyzing optimal disconnected and connected graph designs. To quantify the loss in matching due to constraints on the size, we construct corresponding loss curves for fixed n and k. We conduct a performance analysis between K-chains and K-clusters, which act as proxies for connected and disconnected graph classes respectively. Through these curves, we illustrate a transition in the graph designs from

K-clusters to K-chains as m increases. Our work aims to provide a framework for stakeholders to enhance their supply chain networks' adaptability in an increasingly volatile world.

3 - Environmental impacts of pickup point networks in last mile logistics

Ilkka Leppänen, Aalto University, Espoo, Finland, Lauri Kuula, Juuso Leskinen, Linus Antell

Pickup point networks (PPNs) are being increasingly used by logistics companies in e-commerce. Because PPNs reduce the need for last mile home deliveries, they are often viewed as an environmentally sustainable alternative. However, previous research has not comprehensively taken into account the environmental impacts that incur when customers travel to pickup points. Moreover, the problem is complicated by customers' preferences to chain their parcel pickup trips with other daily commute. Building on the vehicle routing problem, we use operational data and behavioural survey data from a large national postal logistics operator and compare the total environmental impacts that different PPN configurations have, including a configuration with trip chaining. Our results show the importance of accounting for customer habits and preferences in network models that are used to study last mile parcel deliveries.

4 - Correspondent Banking Networks Optimization

NIMA SAFAEI, Scotiabank, Toronto, ON, Canada

Correspondent Banking (CB) Network refers to a network of financial institutions providing cross-border payment services for customers through different channels such as SWIFT, Fedwire, etc. We employ the mathematical programming approach in conjunction with the graph theory to optimize a CB network. Optimizing the network requires decisions to be made to onboard, terminate or restrict the bank relationships to optimize the size and overall risk of the network. This study provides theoretical foundation to detect the components, the removal of which does not affect some key properties of the network such as connectivity and diameter. We find that the correspondent banking networks have a feature we call k-accessibility, which helps to drastically reduce the computational burden required for finding the above mentioned components.

SB78

Regency - 709

Optimization Modeling Software II

Invited Session

Computing Society

Chair: Susanne Heipcke, FICO, Xpress Optimization, Birmingham, United Kingdom

Co-Chair: Bob Fourer, AMPL Optimization Inc., Evanston, IL, 60201, United States

1 - Building and Solving Optimization Problems with the Augmented Api for Fico Xpress Solver

Alexander Biele, FICO, Darmstadt, Germany

The new FICO Xpress Solver API for Java, C# and C++ is designed as an object-oriented layer achieving an optimized interaction between Xpress Solver and model. This leads to a dramatically reduced overall model building time and increased memory efficiency. With the new API all Xpress Solver features are accessible, including full access to the recently added Global Solver. Among the key features of the new Xpress Solver API are the ability to use modern programming concepts such as Collections, Streams, Lambdas, and operator overloading to build expressions and constraints. In the presentation we show that the interface guarantees a consistent user experience across different programming languages. Furthermore, we will provide guidelines on how to efficiently implement optimization models in the new object-oriented API and discuss selected examples.

2 - A modeling language-based approach to automatically recommend first-order optimization methods

Sofiane Tanji, UCLouvain, Louvain-la-Neuve, Belgium, Francois Glineur

Since the advent of modern computational mathematics, the literature on optimization algorithms has been ever-growing, and a wide range of methods are now available. Usually, published algorithms are shown to be applicable to given templates of optimization problems, and are often accompanied by proofs characterizing their convergence rates.

There exist many variations of such templates. Indeed, a template may describe the objective as a single function or a sum of functions with different characteristics (say, smoothness, convexity, existence of a computable proximal operator, etc.), accompanied by constraints that can also be described in various ways (e.g. linear constraints, feasible set with a computable projection operator, functional constraints, etc.).

Hence, for a given instance of an optimization problem, it is not always easy to identify which templates it can match, especially if one wants to consider equivalent reformulations of the problem. This makes the task of choosing an optimization method tedious for the practitioner.

In this talk, we present a modeling language that aims at describing oracle-based optimization problem formulations. Using this language, we propose a framework that automatically checks whether a user-provided optimization problem fits a known template. The framework, implemented as a Python package, relies on an extensive library of optimization methods, described using the above-mentioned language, with their associated known convergence rate for one or several templates, with the goal of automatically ranking applicable optimization methods according to their worst-case theoretical guarantees. The framework also handles some reformulation techniques, allowing more corresponding templates to be potentially identified.

3 - Challenges in Automated Conversion of Optimization Problems

Bob Fourer, AMPL Optimization Inc., Evanston, IL, United States

The range of expressions recognized by modeling languages and solvers has been steadily extended, in ways that make optimization models easier to describe, validate, and maintain — but that make conversion possibilities ever more numerous and complex. Following an introductory survey, this presentation describes a variety of challenges that have been faced in detecting formulations that solvers can handle, and in implementing conversions to forms that solvers require. Issues include distinguishing easy from hard cases, recognizing alternatives preferred by solvers, accounting for side-effects of tolerance settings, and verifying results. Examples are taken from a new solver interface library designed for the AMPL modeling language.

4 - User Friendly Modeling from LINDO Systems

Linus Schrage, LINDO Systems, Inc., Chicago, IL, United States

We describe the latest enhancements to 1) LINDO API, the solver engine, and 2) LINGO, the algebraic modeling language. 3) What'sBest!, the Excel add-in optimizer, Some of the new features are: LINDO API: smarter handling of polynomial functions, and more generally, exotic functions for which derivatives are not available. Substantial improvements to the Global solver, especially linearization, improved reproducibility when using concurrent solver(simplex vs barrier). LINGO: Enhanced editor including model-smart text auto-completion so spelling errors are a thing of the past, and improved performance on large files/models.. What'sBest!: Smart support for more Excel functions, especially in large spreadsheets, plus improved Black Box solver supporting all Excel functions. To help users get started, we also describe our expanded Models Library with smart lookup of various application types: financial models, supply chain, scheduling, production line design, cutting stock, and more.

Sunday, October 20, 12:45 PM - 2:00 PM

SC01

Summit - 320

New Models for Project Management and Scheduling Under Uncertainty

Invited Session

Scheduling and Project Management

Chair: Caihua Chen, Santiago, N/A, United States

Co-Chair: Houcai SHEN, Nanjing University, Nanjing, 210008

1 - An approach for robust resource constrained project scheduling problem

Qian Hu, Nanjing University, Nanjing, China, People's Republic of

Uncertain activity durations and resource disruptions are often found in project scheduling. We study a robust version of the resource constrained project scheduling problem. A robust optimization model, reformulation and a new algorithm are studied. With a computational study, we provide insights about the robustness and analyze the performance of the approach.

2 - Distributed Robust Optimisation Approaches for Resilience and Low Carbon Enhancement of Distributed Power Systems

Caihua Chen, Nanjing University, Nanjing, China, People's Republic of, Houcai SHEN

Resilience and Low Carbon are two trend to quantity the ability of distribution power system. In this work, we discuss the approach which accesses to microgrids in the main grid, which can handle significant power-disrupting incidents due to their islanding capabilities and the potential to support renewable energy penetration. In terms of the definition of resilience, we propose a multi-stage distributionally robust optimization model which capture the preparation, resoinse, and restoration three states. We further add the consideration of the cost of carbon emissions in our objection. The proposed framework stipulates that the chance constraints hold under the worst-case distribution within a novel ambiguity set, which incorporates the Wasserstein distance and the first-order moment. To solve the issue, the model is reduced to a mixed-integer linear programming problem that can readily be implemented. Numerical experiments are carried out on the IEEE 34-bus test systems to show the significantly enhancement of the addtion of microgrids.

3 - R&D Project Investment with Delay Option and Information Spillover Effect

Houcai SHEN, Nanjing University, Nanjing, China, People's Republic of, Yu Ge

This study explores how two competing companies make investment decisions in R&D markets, considering positive policy uncertainty and negative technology uncertainty. Additionally, this paper demonstrates that the delay option may bring profits for the investors from information spillover of the pioneer firm and from the waiting value of the market. Since different firms may have different information spillover effect on the followers, the optimal investment decisions of firms are also considered under asymmetric information spillover effects in this paper. Game theory is also employed to obtain the Nash equilibrium investment policies for these two competing firms.

4 - Dynamic Assortment with Online Learning under Threshold Multinomial Logit Model

Wenxiang Chen, Nanjing University, Nanjing, China, People's Republic of, Caihua Chen, Houcai SHEN, Ruxian Wang, Weili Xue

Consumers often find themselves overwhelmed by extensive assortments offered by online retailers and show bounded rationality behavior. However, existing literature on assortment optimization didn’t consider consumers' such bounded rationality behavior. This motivates us to employ a simple and effective two-stage consider-then-choose model, namely the Threshold Multinomial Logit (TMNL) model to investigate the online assortment optimization problem. The TMNL model characterizes consumers' endogenous consideration sets formation by the threshold effect. This endogenous dependency can capture more flexible substitution patterns than the classical MNL choice model, but it also creates great difficulties for online learning. In the offline assortment setting, we analyze the properties of optimal assortment and propose an efficient assortment optimization algorithm outperforms the benchmark. In the online setting with unknown customer preferences and consideration set formation, we propose online learning algorithms that achieve nearly optimal regret bounds under both instance-independent and instance-dependent conditions. To the best of our knowledge, we are the first work to consider online assortment problem with consumers' endogenous consider-then-choose behavior. Moreover, our algorithm is extended to the contextual learning setting, effectively mitigating the impact of the number of products on its performance. Extensive numerical experiments validate the efficacy of our proposed algorithms.

SC02

Summit - 321

Journal of Quality Technology Invited Session

Award Session

Quality, Statistics and Reliability

Chair: Rong Pan, Arizona State University, Tempe, AZ, United States

1 - Directional fault classification for correlated high-dimensional data streams using hidden Markov models

Modern manufacturing systems are often installed with sensor networks which generate high-dimensional data at high velocity. These data streams offer valuable information about the industrial system's real-time performance. If a shift occurs in the manufacturing process, fault diagnosis based on the data streams becomes a fundamental task as it identifies the affected data streams and provides insights into the root cause. Existing fault diagnostic methods either ignore the correlation between different streams or fail to determine the shift directions. In this paper, we propose a directional fault classification procedure that incorporates the between-stream correlations. We suggest a three-state hidden Markov model that captures the correlation structure and enables inference about the shift direction. We show that our procedure is optimal in the sense that it minimizes the expected number of false discoveries while controlling the proportion of missed signals at a desired level. We also propose a deconvolution-expectation-maximization (DEM) algorithm for estimating the model parameters and establish the asymptotic optimality for the data-driven version of our procedure. Numerical comparisons with an existing approach and an application to a semiconductor production study show that the proposed procedure works well in practice.

2 - Spatio-temporal process monitoring using exponentially weighted spatial LASSO

Peihua Qiu, University of Florida - Department of Biostatistics, Gainesville, FL, United States, Kai Yang

Spatio-temporal process monitoring (STPM) has received a considerable attention recently due to its broad applications in environment monitoring, disease surveillance, streaming image processing, and more. Because spatio-temporal data often have complicated structure, including latent spatio-temporal data correlation, complex spatio-temporal mean structure, and nonparametric data distribution, STPM is a challenging research problem. In practice, if a spatio-temporal process has a distributional shift (e.g., mean shift) started at a specific time point, then the spatial locations with the shift are usually clustered in small regions. This kind of spatial feature of the shift has not been considered in the existing STPM literature yet. In this paper, we develop a new STPM method that takes into account the spatial feature of the shift in its construction. The new method combines the ideas of exponentially weighted moving average in the temporal domain for online process monitoring and spatial LASSO in the spatial domain for accommodating the spatial feature of a future shift. It can also accommodate the complicated spatio-temporal data structure well. Both simulation studies and a real-data application show that it can provide a reliable and effective tool for different STPM applications.

3 - Quality prediction using functional linear regression with in-situ image and functional sensor data

Jionghua Jin, University of Michigan, Ann Arbor, MI, United States, Wenbo Sun

This article studies a general regression model for a scalar quality response with mixed types of process predictors including process images, functional sensing signals, and scalar process setup attributes. To represent a set of time-dependent process images, a third-order tensor is employed for preserving not only the spatial correlation of pixels within one image but also the temporal dependency among a sequence of images. Although there exist some papers dealing with either tensorial or functional regression, there is little research to thoroughly study a regression model consisting of both tensorial and functional predictors. For simplicity, the presented regression model is called functional linear regression with tensorial and functional predictor (FLR-TFP). The advantage of the presented FLR-TFP model, which is compared to the classical stack-up strategy, is that FLR-TFP can handle both tensorial and functional predictors without destroying the data correlation structure. To estimate an FLR-TFP model, this article presents a new alternating Elastic Net (AEN) estimation algorithm, in which the problem is reformed as three sub-problems by iteratively estimating each group of tensorial, functional, and scalar parameters. To execute the proposed AEN algorithm, a systematic approach is developed to effectively determine the initial running sequence among three sub-problems. The performance of the FLR-TFP model is evaluated using simulations and a real-world case study of friction stir blind riveting process.

4 - Directional fault classification for correlated high-dimensional data streams using hidden Markov models

Yan He, East China Normal University, Shanghai, China, People's Republic of

Modern manufacturing systems are often installed with sensor networks which generate high-dimensional data at high velocity. These data streams offer valuable information about the industrial system's real-time performance. If a shift occurs in the manufacturing process, fault diagnosis based on the data streams becomes a fundamental task as it identifies the affected data streams and provides insights into the root cause. Existing fault diagnostic methods either ignore the correlation between different streams or fail to determine the shift directions. In this paper, we propose a directional fault classification procedure that incorporates the between-stream correlations. We suggest a three-state hidden Markov model that captures the correlation structure and enables inference about the shift direction. We show that our procedure is optimal in the sense that it minimizes the expected number of false discoveries while controlling the proportion of missed signals at a desired level. We also propose a deconvolution-expectation-maximization (DEM) algorithm for estimating the model parameters and establish the asymptotic optimality for the data-driven version of our procedure. Numerical comparisons with an existing approach and an application to a semiconductor production study show that the proposed procedure works well in practice.

SC03

Summit - 322

From Data to Decisions: Advancing Smart Healthcare through Machine Learning

Invited Session

Quality, Statistics and Reliability

Chair: Jianxin Xie, University of Virginia, Charlottesville, VA, United States

Co-Chair: Zekai Wang, Fairfield University, Fairfield, CT, United States

1 - Generating Counterfactual Explanations in Breast Cancer Prediction

Siqiong Zhou, Arizona State University, Tempe, AZ, United States, Ashif Iquebal

Imaging phenotypes extracted via radiomics of magnetic resonance imaging have shown great potential in predicting the treatment response in breast cancer patients after administering neoadjuvant systemic therapy (NST). Understanding the causal relationships between the treatment response and Imaging phenotypes, Clinical information, and Molecular (ICM) features are critical in guiding treatment strategies and management plans. Counterfactual explanations provide an interpretable approach to generating causal inference. However, existing approaches are either computationally prohibitive for high dimensional problems, generate unrealistic counterfactuals, or confound the effects of causal features by changing multiple features simultaneously. In previous work, we proposed Sparse Counterfactual Generative Adversarial Networks (SCGAN) for generating counterfactual instances to reveal causal relationships between ICM features and the treatment response after NST. SCGAN generates sparse and diverse counterfactual instances that also achieve plausibility and feasibility. In this work, we improve SCGAN on coefficient tuning and enhance model efficiency, making it a valuable tool for understanding the causal relationships between ICM features and treatment response.

2 - Generative Model and Incremental Learning for Longitudinal Electric Health Record Data

Ziyang Zhang, Oklahoma State University, Stillwater, OK, United States, Chenang Liu

The combination of generative models and incremental learning offers a powerful approach for handling continuously evolving data while generating realistic synthetic Electronic Health Records (EHR) samples for training and research. This work explores the integration of these techniques to improve predictive models and patient outcomes, focusing on challenges such as longitudinal data handling, data incompleteness, and model adaptability for dynamic non-equally spaced longitudinal data.

3 - Dual-Output Deep Learning for Reducing Unnecessary Lab Tests in ICU

Nasim sadat Mousavi, University Of Arkansas, fayetteville, AR, United States, Shengfan Zhang

Unnecessary laboratory testing in Intensive Care Units (ICUs) imposes excessive costs with minimal clinical benefits for patients, particularly for those undergoing Gabapentin therapy. The primary goal of this research is to develop a dual-output Long Short-Term Memory (LSTM) neural network model capable of predicting laboratory test results for ICU patients on Gabapentin, using the MIMIC-IV data. Encompassing demographic information, vital signs, and historical lab results for critical biomarkers, the LSTM model can help reduce unnecessary testing and optimize healthcare resource utilization. The primary output of this model accurately forecasts the quantitative values of future laboratory tests. Simultaneously, the secondary output evaluates the probability of the need for each test based on a novel loss function that balances the accuracy of prediction and the efficiency of test reduction. This loss function can enhance the decision-making process by minimizing the number of performed tests and prediction errors, aiming for a reduced number of lab tests without compromising diagnostic accuracy. Decision thresholds derived from these probabilities determine whether tests should be performed, minimizing redundant testing, and aligning with clinical needs. This approach introduces a systematic shift towards precision medicine by tailoring lab testing to patients' needs. It promises to enhance the efficiency of lab testing in ICU settings by reducing unnecessary procedures, optimizing the use of medical resources, and enhancing patient care.

Keywords: LSTM, laboratory test reduction, Gabapentin, Dynamic decision-making

SC04

Summit - 323

Computation Service and Data Quality in Manufacturing Industrial Internet

Invited Session

Quality, Statistics and Reliability

Chair: Yingyan Zeng, University of Cincinnati, Blacksburg, VA, United States

Co-Chair: Ran Jin, Virginia Tech, Blacksburg, VA, United States

1 - Ontology-guided Data Sharing with Differential Privacy in Interconnected Additive Manufacturing

Hui Wang, Florida A&M University-Florida State University College of Engineering, Tallahassee, FL, United States

The scarcity of measured data crucial for defect identification often presents a challenge to additive manufacturing process development and certification. Knowledge sharing has become an emerging solution to the problem. However, manufacturers are concerned about privacy loss during knowledge sharing. This study integrates a differential privacy model with federated learning to improve knowledge sharing with privacy protection, thereby using the knowledge sharing to improve defect identification given limited data. The framework addresses the following issues, which existing federated learning fails to deal with (1) How to represent information in the knowledge to be shared; (2) How to select and mask the shared knowledge while balancing the trade-off between privacy protection and defect identification; and (3) How to quantitatively control privacy loss.

2 - A fast nonparametric process monitoring scheme in fully decentralized systems

Jiahui Zhang, University of Wisconsin-Madison, Madison, WI, United States

Along with the development of big-data technologies, multi-sensor systems have become more and more prevalent in manufacturing industries. To ensure the production process is in normal condition, process monitoring algorithms have been applied to multi-sensor systems, in which a central server aggregates information from all sensors and make decisions about the process condition. This kind of centralized systems, however, highly relies on the central server and cannot protect the privacy of the sensor data. This leads to the need of decentralized

systems, in which each edge device can communicate with each other and individually raise the system alarm. The goal of this paper is to develop a fully decentralized process monitoring scheme for fast online change detection. Our monitoring framework includes the design of a nonparametric device-level statistic, a smart strategy for inter-device communication, and an alarm-raising mechanism which responds quickly to mean shifts and is resilient to partial function failures of the devices. The proposed method enjoys good theoretical properties which enables its effectiveness and efficiency. The outperformance of the proposed method is shown in numerical experiments and a case study.

3 - Memory Span Reinforcement Learning: Bridging Short-Term Actions and Long-Term Goals in Dual-Agent Systems

Zehua Dong, University at Buffalo , Buffalo , NY, United States, Xiaoyu Chen, Xiaomei Wang

Recent advancements in multi-agent reinforcement learning (MARL) systems have made significant contributions to a variety of fields, including autonomous vehicle coordination, financial trading, and grid management. These systems require the cooperation among multiple agents, enabling dynamic interactions in real-world applications. However, the lack of clearly defined roles in current multi-agent reinforcement learning (MARL) systems prevents agents from effectively fulfilling their own goals, necessitating methodologies like Centralized Training with Decentralized Execution (CTDE) and fully decentralized training. These approaches are essential for coordinated learning and independent execution, but they limit the transferability from a well-perceived environment to similar-but-not-identical environments.

Motivated by underdefined cooperation, we propose a novel cooperative dual-agent reinforcement learning framework that mimics human cognitive processes. It enables inter-agent coordination by assigning different agents with different but cooperative goals. Inspired by human information processing (HIP) model, we created a hierarchical structure featuring two agents: a long-term agent and a short-term agent. This architecture mirrors the HIP model where short-term actions are driven by immediate stimuli (bottom-up processes), represented by immediate rewards; while long-term strategies are guided by overarching goals and knowledge (top-down processes), represented by historical trajectories. The long-term agent strategically oversees the learning process and directly controls the action space of the short-term agent, thus enhancing learning efficiency and transferability across similar-but-not-identical environments. Our framework is particularly suitable for applications requiring domain adaptation.

4 - On High-quality Data Trade

Ran Jin, Virginia Tech, Blacksburg, VA, United States

The advancement of artificial intelligence (AI) models heavily relies on the availability of large, informative, and high-quality datasets. However, in critical applications such as manufacturing and transportation, it takes a long time and intensive effort for high-quality data collection and preparation for AI training and deployment. In this paper, we will introduce a few methodologies to define dataset quality, high-quality data generation, and performance-oriented data trade with privacy preserving consideration. A representation learning framework in Directed Acyclic Graph Neural Network is introduced. The effectiveness of the work is validated by two case studies: a semiconducting manufacturing network and design and manufacturing network of Microbial Fuel Cells.

SC05

Summit - 324

DEI Ambassador Program

Flash Session

Diversity, Equity, and Inclusion

Chair: Daniel Reich, Naval Postgraduate School, Monterey, CA, United States

Co-Chair: Banafsheh Behzad, California State University, Long Beach, Long Beach, CA, United States

Co-Chair: David Czerwinski, San Jose State University, San Jose, CA, United States

1 - Informs K-12 Education Outreach and Networking Program

Neil Desnoyers, Saint Joseph's University, Philadelphia, PA, United States, Zihan Zhang, Fenglian Pan

The primary goals of the K-12 Education Outreach and Networking Program 2024 DEI Ambassador Project are to extend the footprint of and strengthen INFORMS' K-12 outreach program. The major components of our work this year are:

1. Organize an in-person focus group for secondary math teachers in the Philadelphia region, thereby extending the footprint of INFORMS' K-12 outreach program.
2. Strengthen INFORMS' K-12 outreach program by holding a series of webinars introducing K-12 outreach to the INFORMS membership and others.
3. Facilitate communication and collaboration among the K-12 community within INFORMS by organizing the second K-12 panel at the 2024 INFORMS Annual Meeting.

An update on the project will be provided.

2 - Uncovering Racial Bias in Automated Traffic Law Enforcement: Extensions

Chrysafis Vogiatzis, University of Illinois Urbana-Champaign, Urbana, IL, United States, Chris Raymond-Bertrand, Eleftheria Kontou

In 2023, two of the classes at the University of Illinois Urbana-Champaign adopted a novel culturally relevant case study that focused on a contemporary societal issue: injustice in automated traffic law enforcement. The case study, which received the INFORMS Best Case Runner-up recognition in the INFORMS Annual Meeting in 2023, had two main components. The first one focused on proper statistical hypothesis testing using real-life demographic and traffic data. The second one introduced a new network optimization problem in the form of a side-constrained shortest path. Students that were introduced to this case study found it very interesting, specifically due to the connection and implications that operations research and the management sciences have for policy. In this extension, we study different origin-destination pairs in the city of Chicago. That way, we can establish how automated traffic law enforcement disproportionately hurts travelers and commuters from different parts of the city (with different socioeconomic characteristics) when trying to reach specific points of economic and cultural activity.

3 - Data Analysis&Mdash;Perspectives of DEI Initiatives in Academia & Industry

Mubarak Iddrisu, University of Massachusetts Boston, Boston, MA, United States, Ogechi Vivian Nwadiaru, Hamid Arzani, Andrea Hupman, Allison Reilly, Jun Zhuang, Gul Kremer

This study examines the alignment between organizational Diversity, Equity, and Inclusion (DEI) commitments at the management level and individual employee perceptions in practice. Organizational commitment to DEI initiatives is pivotal for fostering inclusive workplace cultures. However, a gap could exist between aspirational ideals and actual behaviors, raising questions about the practical implementation of DEI values within organizations. Drawing from existing literature and frameworks, including McKinsey & Company's "Diversity Wins: How Inclusion Matters," we address this gap by investigating the specific barriers and challenges hindering the translation of DEI commitments into action. Our research questions include; How do DEI commitments at the management level align with individuals' perceptions in practice? How do organizations manifest their DEI value statements into actions? How to design metrics to measure DEI initiatives?

Together these questions explore how DEI commitments at the management level align with employees' day-to-day experiences, shedding light on the effectiveness of DEI initiatives within organizations. Through survey-based methodology, we analyze perceptions from both managerial and non-managerial roles, focusing on aspects such as inclusion efforts, alignment with institutional DEI statements, and perceptions of equity. By probing into these dimensions, our study provides evidence-based insights to support organizations, especially in the operations research and management sciences striving to create genuinely inclusive cultures.

4 - Unite: Uplifting Networks for Inclusive Transformation and Equity

Fatemeh Nosrat, Rice University, Houston, TX, United States, Aysenur Karagoz

We intend to implement tailored mentorship programs, workshops, and resources to address the unique challenges faced by to address the unique challenges faced by women, African Americans, Asians, Middle Eastern individuals, Hispanic and Latino Americans, Native Americans, individuals with disabilities, and members of the LGBT+ community within the INFORMS community. Through collaboration with existing minority communities, such as WORMS and Pride Forum, we aim to foster partnerships and amplify diverse voices. Our initiatives focus on creating a more inclusive environment, attracting a broader range of students, and celebrating diversity through events and success stories. By incorporating feedback mechanisms and regular assessments, we are committed to continuous improvement and ensuring the success of our DEI efforts.

5 - Active Engagement with DEI Students from Data-Related Undergraduate Programs in the INFORMS Data Mining Society and Beyond

Ying Lin, University of Houston, Houston, TX, United States, Andi Wang, Chun-An Chou, Nathan Gaw

Data science (DS) has drawn increasing attention in academia and every conceivable industry. As a result, there is ever-increasing motivation for students of diverse backgrounds to learn data mining and decision analytics (DMDA) skillsets for their future career. However, it brings apparent difficulties and barriers to train and foster students who were not in DS-related majors. This project is a continuation of the project last year titled, Active Engagement with DEI Students from Data-Related Master's Programs in the INFORMS Data Mining Society and Beyond. Due to the massive success in this project from last year, we chose to continue the direction of these efforts by (1) engaging with DEI students in undergraduate programs, (2) gaining insights on how we can better engage with minority K-12 populations in STEM topics, and (3) initiating with universities that have underserved Hispanic demographic. To achieve the goals, we initiate the first data mining undergraduate poster session, host a data competition with K-12 focus, organize a DEI-oriented DMDA workshop and evaluate the impact on DEI students post the DMDA workshop.

6 - Operations Research in Women's Health and Healthcare: a Scoping-Inspired Narrative Review and Roadmap for the Future

Rachel Stephenson, University of Toronto, Toronto, ON, Canada, Rachel Wong

Operations Research (OR) approaches have been applied to derive meaningful policy insights and effective solutions to a broad range of healthcare problems, but we believe that there is a gap in OR literature addressing issues that differently or disproportionately impact women's health and healthcare (hereafter referred to as "issues in women's health and healthcare" for brevity). As such, we are undertaking a project to identify this gap and promote a research agenda to close it. The project consists of three phases:

Phase 1: Perform a scoping-inspired review of literature on significant issues in women's health and healthcare studied and published in top OR journals over the past 10 years.

Phase 2: Distribute a survey to experts in women's health and healthcare with the research gaps identified in Phase 1. Experts will be asked to prioritize research areas, identify specific subproblems within broad areas that may be solved using OR approaches, and identify any emerging issues that were excluded from the list provided.

Phase 3: The results from Phases 1 and 2 will be synthesized and disseminated in a survey to the OR community. Experts will be asked to describe how OR could be applied to solve these problems.

The primary outcome of this work will be the identification of opportunities for novel applications of OR in women's health and healthcare. The goal of this project is to promote applications of OR to women's health and healthcare issues, and therefore, to ultimately improve women's health outcomes.

7 - Innovation In Operations Research: Fostering Equity and Justice In Problem Formulation

Sadan Kulturel-Konak, Penn State Berks, Reading, PA, United States, Abdullah Konak, Sara Abu Aridah

The fields of Operations Research (OR) and Management Science (MS) have traditionally prioritized cost, benefit, and efficiency in problem-solving. While these considerations are undoubtedly vital, our proposed project seeks to push the boundaries of traditional thinking by introducing new dimensions – justice and equity- in problem-solving. By integrating case studies that highlight social justice and equity aspects, we aim to broaden the scope of discussions within introductory-level OR and MS courses. Our emphasis extends beyond merely solving problems efficiently by only focusing on conventional cost/profit optimization; we aspire to actively contribute to the cultivation of a framework that also prioritizes and fosters principles of equity and justice; by doing so, we hope to provide practices for students to nurture an 'equity mindset.' Integrating social justice and equity perspectives into OR and MS with a multi-objective approach is essential for several reasons. Firstly, it aligns with the dynamic evolution of societal values, where ethical considerations are increasingly taking center stage. Secondly, by instilling these concepts at the foundational level, we are not only equipping future professionals to tackle real-world challenges but also fostering a more inclusive and diverse community within the fields of OR and MS. Our approach involves the development of problems and case studies that illustrate the intersection of OR with equity and social justice principles. We plan to focus on the basic and introductory level problems that are typically used in the first OR and MS courses.We will also create instructor companions summarizing pedagogical strategies to support instructors.

8 - Techfest Dayton: Large-Scale K-12 Stem Outreach Targeting DEI-Focused Organizations, Presenters, & Communities

Kara Combs, Air Force Research Laboratory, Wright-Patterson AFB, OH, United States, Nathan Gaw, Trevor Bihl

TechFest Dayton is the premiere K-12 Science, Technology, Engineering, and Mathematics (STEM) Outreach event held over two days in Dayton, OH. Dayton faces several socio-economic challenges that affect the city's poverty, substance abuse, and education levels. Hosted by its namesake 501(c)3 foundation, TechFest Dayton aims to improve the statistics of these areas by exciting local kids about STEM and higher education through accessible hands-on activities and demonstrations. In 2024, over 1,100 kids and an estimated 2,500+ attendees came to TechFest Dayton. With 2023 attendee feedback focused on greater consideration for DEI-related issues, the 2024 committee was happy to triple the number of DEI-focused exhibitors, schedule presenters such that 20% were female, and add a sensory-friendly hour. This is in addition to ensuring the event remains free for all participants – attendees, exhibitors, and presenters. Of the children in attendance 42.5% of them belonged to female- and minority-identifying populations and over half were first-time attendees at the event. The local Cincinnati-Dayton INFORMS chapter co-exhibited with the Institute of Industrial and Systems Engineers Dayton-Cincinnati professional chapter and provided an introduction to operations research and engineering for the kids in attendance. Their joint booth had multiple activities including a getting ready for school activity (management/planning) and two computerized games aimed to teach the kids about statistics and programming. In conclusion, the DEI Ambassador grant program provided crucial financial support for TechFest Dayton and also played a pivotal role in diversifying and raising awareness about operations research within the Dayton community.

9 - Diverse Voices in Operations Research and Management Sciences: a Storytelling Platform

Will Cong, Cornell University, Ithaca, NY, United States, Jihye Jang, Yutong Meng, Luyao Zhang

<div overflow-hidden"=""><div h-full"="">

<div text-token-text-primary"="" dir="auto" data-testid="conversation-turn-3" data-scroll-anchor="true">

This project aims to create an inclusive online platform showcasing the diversity within the INFORMS community, fostering interdisciplinary collaboration, and amplifying underrepresented voices. Through a series of activities and deliverables, including email Q&A interviews, eCornell webinars, and interview article publications, the project highlights contributions from academia, industry, and policy circles. These efforts are designed to enhance engagement and promote diversity and inclusion across various spheres. Interview questions tailored for scholars, practitioners, and policymakers delve into research impact, practical applications, policy implications, challenges, and future directions, emphasizing the importance of interdisciplinary collaboration in addressing contemporary challenges in operations research and management sciences.

SC06

Summit - 325

Facility/Hub Location 2

Invited Session

Location Analysis

Chair: Gita Taherkhani, Loyola University Chicago, Chicago, IL, United States

1 - Towards Classifying Instances for the Capacitated Facility Location Problem

Stefan Nickel, Karlsruhe Institute of Technology, Karlsruhe, Germany, Hannah Bakker

The Capacitated Facility Location Problem (CFLP) is a core problem of location. As such, it attracted much research presenting novel solution methods that are typically evaluated with an empirical assessment of a chosen set of benchmark instances. An algorithm is considered successful if it outperforms other algorithms on average. Such an average-case analysis comes with limitations as it fails to incentivize a thorough discussion of an algorithm's strengths and weaknesses, and bears a risk of inherent bias in the selected instances. The latter risk is particularly prevalent for the CFLP for which only very few real-world-based test instances exist, and a significant proportion of the available instances are based on the same generation procedure. We perform a systematic analysis of the features that make a CFLP instance hard or easy for a particular algorithm using Instance Space Analysis (ISA), a recent framework that supports objective testing and assessing the diversity of test instances. While several works exist that propose a variety of characterizing features for different combinatorial optimization problems, the CFLP differs from these problems in that it contains a more diverse set of parameters that are even measured on different scales (costs and volume) and have partially overlaying effects on the solution structure. We identify hardness revealing features and link them to the performance of a greedy heuristic, Kernel Search, and a commercial MIP solver with different parameter settings.

2 - Strategic Expansion of Freight Transportation Hub Networks Under Uncertainty

Gita Taherkhani, Loyola University, Chicago, IL, United States, Hao Li, Sibel Alumur Alev, Mike Hewitt

We focus on freight transportation carriers that transport shipments that are small relative to vehicle capacity and incur transportation costs that exhibit economies of scale. To achieve profitability, such carriers route shipments through a network of hubs. The locations of a carrier's hubs dictate the customer markets it can profitably serve. We consider a carrier that seeks to increase its profitability by growing its customer base. To do so, it seeks to expand its network into new regions by merging with carriers that already operate in those regions. We focus on how, and by how much, the network that results from such a merger should be redesigned to maximize profitability. To study this issue we present deterministic and stochastic profit-maximizing hub location models that capture the interaction between economies of scale and profitability. The models differ in whether and how they model the recognition of uncertainty in shipment sizes in the decision-making process. We perform a case study based on operations from a multi-regional less-than-truckload freight transportation carrier to derive insights into the profitability of different redesign strategies. We analyze the networks prescribed by solutions to these optimization models to derive insights into how a network that results from a merger should be redesigned. We also study how uncertainty impacts the structure of the redesigned networks.

3 - Periodic Vehicle Location Routing and Scheduling in Humanitarian Logistics

Bahar Yetis Kara, Ihsan Dogramaci Bilkent University, Ankara, Turkey, Cagla Dursunoglu, Okan Arslan

In the recovery phase of a humanitarian crisis, periodic demand service is an important aspect of mobile services, referring to the requirement for a service that occurs at regular intervals. Mobile barbers, mobile laundry, and mobile showers are examples of mobile services that are in demand on a regular basis. The demand occurs at certain periods since the impacted population may frequently need such services. In order to service this demand type, the customer locations need to be visited at certain intervals. This creates a regular pattern of demand that the mobile service provider must satisfy by providing flexible scheduling and efficient service delivery. Each service type has its own set of mobile service units and each demand point should be served at least a certain number of times by the various mobile services within the specified planning horizon. The initial and terminal locations for vehicles and their routes should be determined, which collectively induces the order of demand visits. Furthermore, the duration of mobile service visits at a demand location is also a decision. There may be incompatible services that are forbidden to visit a location simultaneously. Additionally, the visits by mobile service centers offering the same services should be spaced apart by at least a certain duration of time.

SC07

Summit - 327

Innovative Applications in Business and Technology

Flash Session

Flash

Chair: Molly Goldstein, University of Illinois, Urbana-Champaign, 104 S. Mathews Ave., Urbana, IL, 61801, United States

1 - Craftsmanship versus Investment: Exploring Factors Affecting the Nonfungible Token (NFT) Projects Performance

Yanxin Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Jingzhao An, Xiaoni Lu, Xi Zhao

The nonfungible token (NFT) market has experienced tremendous volatility in recent years, prompting a need for deeper investigation into the effective design of NFT projects amid this period of uncertainty. This study therefore investigates how NFT design-related features influence the performance of NFT projects in the short- and long-term. From the perspective of consumption, this study distinguishes NFT consumption into competency-based and investment-based consumption, according to the dual attributes of information goods and tokens associated with NFTs. Correspondingly, we identify two types of design features of NFT collections, namely, competency-related features (i.e., NFT image complexity and consistency), as well as investment-related features (i.e., initial prices and royalties). Using transaction data of 3,037 NFT projects on Ethereum blockchain, we find that NFT image complexity has a significant inverted U-shaped effect on long-term performance due to the possible information overload. Image consistency within NFT projects has positive impacts on both short-term and long-term performance due to the formulation of brand symbolism. Royalty as the resale cost negatively affects the short-term performance. Additionally, royalty and initial price exhibit inverted U-shaped impacts on long-term performance, owing to their mixed roles as both costs and signals. As for the moderating effect of market uncertainty, time pressure brought by uncertainty enhances the impact of image complexity and royalty, while weakening the role of initial price as a quality signal. Our findings provide both theoretical and practical implications for NFT project design, enhancing the understanding of consumer behavior in the Web3 context.

2 - The Impact of Social Effect on Funding Success

Sanghyun Park, Baruch College, City University of New York, New York, NY, United States, Qiang Gao

The exponential growth of online donation-based crowdfunding for education has raised questions about the influence of social effects on funding success and the justification behind backers' decisions. This pioneering study extends beyond conventional crowdfunding research by examining the unique dynamics of donation-based crowdfunding. Utilizing a comprehensive dataset from DonorsChoose.org and the U.S.

Department of Education, the analysis reveals an inverted U-shaped relationship between predicted social effects and funding success, indicating that projects with moderate social effects are more likely to succeed compared to those with low or high appeals. The study also investigates post-donation interactions between fundraisers and backers, providing insights into the reasoning behind backers' decisions. The results demonstrate that social effects reflected by fundraisers's unobservable attribute, such as dedication, have a curvilinear relationship with funding impact. Robustness tests and checks for endogeneity confirm the consistency of the findings. This research highlights the critical need for strategically balanced social effects in educational crowdfunding narratives to optimize funding success. It significantly contributes to the broader crowdfunding literature by elucidating the distinct behaviors of donors in the non-return, charitable context of educational fundraising.

3 - Using Simulation to Enhance Customer Experience and Reduce Congestion Within a Non-Profit Food Bank

Meredith DePuy, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, Kara Gossard, Madison Harmon, Kimberly Ellis

A non-profit food bank, sponsored by more than 25 local churches, provides service to 300 families (approximately 900 people) monthly in southwest Virginia. The food bank operates four days a week (for 1.5 hours each day) and is fully staffed by volunteers, with no paid employees. With increasing demand and rotating volunteers, the pantry needs to improve and standardize processes to enhance customer experience. For this project, we evaluate multiple layout, process, and flow alternatives using discrete event simulation. The results are used to streamline processes and justify capital investments to improve customer flow and overall experience. In addition, the results also provide valuable insights for similar food banks.

4 - Managing cross-departmental knowledge flows to accelerate learning in a high-tech firm

Alex Alblas, Eindhoven University of Technology , Eindhoven, Netherlands, Massimo Manca, Fred Langerak

This study evaluates the differential effect of cross-departmental knowledge flows on organizational learning and operational performance in a firm operating in the semiconductor industry. Knowledge transfers across departments via several mechanisms, which have different returns and require different resources. Previous research examined knowledge transfer mechanisms between departments (R&D, production, service) separately. Our study integrally estimates knowledge flows' differential effect on learning and performance simultaneously. Using a rich panel dataset concerning knowledge flows between departments and service performance, we analyze how the service department learns to improve machine availability via each knowledge flow. Knowledge flows are quantified with the cumulative number of knowledge transfer tasks spanning departmental boundaries, such as problem-solving tasks, work instructions, and design changes. Our study contributes by shedding light on the most effective cross-departmental knowledge flows to accelerate the learning curve in the context of a knowledge-intensive firm.

5 - Traditional vs. Generative Design: Student Artifacts and Decisions

Molly Goldstein, University of Illinois, Urbana-Champaign, Urbana, IL, United States

The rise of generative tools powered by AI has created an urgent desire in many industries to capitalize on the powerful capabilities of these tools, but an understanding of how these tools are utilized and perceived by students is lacking. Generative Design is an AI tool in the computer-aided design software Fusion 360 that can be used to accelerate the design-to-make process. Generative Design achieves this by taking user-defined constraints, existing assembly geometry, and obstacle geometry to generate numerous potential design solutions that have been iterated upon one another. The designer can then select which model to continue developing based on several calculated metrics. However, since generative design produces a range of possible solutions, students need to be able to articulate why those chose a particular solution.

Research Questions: How do generative design tools affect the quality of engineering student's design artifacts and their design decisions?

An observational study was conducted with engineering student participants of varying levels of standing (undergraduate and graduate) and CAD experience. Students were tasked with completing two separate challenges in which they had to design and model a structural component for an existing assembly with traditional parametric CAD techniques and with generative design tools. The collected design artifacts were assessed for quality using the Trade-off Value Protocol and the Informed Design Teaching and Learning Matrix. We compared student rationale for their design process in both the traditional and generative design environments. Preliminary results indicate that students differ on their design rationale in the two environments.

6 - Investigating the impact of behavior representation on targeted ads effectiveness: An empirical study

Wei Xiong, Mercer University, Atlanta, GA, United States

The success of behavioral targeting (BT) hinges largely on how effectively online users are classified, as advertisers prioritize reaching the most relevant audience for their ads. This study considers BT as a user classification problem and describes a machine learning-based solution. We introduce three user behavior representation methods and compare them empirically using the area under the receiver operating characteristic curve (AUC). The experimental results confirm the effectiveness of BT on user classification and provide a validation of BT for online advertising. Notably, combining user search queries, clicked URLs, and clicked ads yields the most significant performance improvement compared to a baseline of user location. Furthermore, this study also investigates the temporal dimension of user behavior by analyzing how historical data length influences targeting performance.

SC08

Summit - 328

Optimization Applications in Defense

Invited Session

Military and Security

Chair: Robert Curry, University of Arkansas, Fayetteville, AR, United States

1 - Mixed-integer Programming Models for solving Network Flow Optimization Problems in Contested Settings

Robert Curry, University of Arkansas, Fayetteville, AR, United States

Various applications require dynamic flow solutions characterized by a schedule of network flows consecutively transmitted over a sequence of successive periods. For these schedules, we assume flows transmit via arcs *during* periods while flows *reside* at nodes from one period to the next. Within this context, we introduce the Maximum Value Dynamic Network Flow Problem (MVDFP) in which we seek to maximize the cumulative *value* of a non-simultaneous network flow schedule that accumulates node *value* whenever some minimum amount of flow resides at a node between periods. For solving the MVDFP, we first introduce a large mixed-integer program (MIP). As this MIP can become computationally-expensive for large networks, we also present a computationally-effective heuristic that sequentially solves a series of smaller, more manageable MIPs. This heuristic approach determines a high-quality solution significantly faster than the MIP obtains an optimal solution by dividing the full network flow schedule into a sequence of consecutive shorter network flow subschedules. In many cases this approach produces an optimal solution in a fraction of the MIP's computational time. We present extensive computational results to highlight our heuristic's efficacy and discuss future research avenues.

2 - Approximate Linear Model for Optimal Surveillance Deployment

Daphne Skipper, United States Naval Academy, Annapolis, MD, United States, Sebastian Martin, Anna Svirsko, Esra Toy

We consider the problem of optimally selecting and locating surveillance devices to detect stealth traffic transiting an operational theater. These surveillance decisions are based on historical traffic data, as well as sensor ranges and detection capabilities. Accounting for overlapping sensor coverage leads to a mixed-integer non-convex optimization model. A log transformation converts the model’s nonlinearity into convex univariate constraints. By careful selection of linearization points, we develop an approximate linear model that is within a specified error tolerance. We solve this problem in the context of detecting illegal fishing off the coast of Thailand to demonstrate the viability of the approximate linear model.

3 - Optimal Multi-Searcher Maneuver with Uptime / Downtime Constraints

Anna Svirsko, United States Naval Academy, Annapolis, MD, United States, Sebastian Martin, Miles Dixon, Daphne Skipper, Esra Buyuktahtakin Toy

When surveillance devices detect a stealth target that is transiting through an operational theater, an area of uncertainty (where the target may be located) is generated and continues to grow until the target is precisely located. We consider the problem of optimally routing search assets to find and apprehend targets following the generation of one or more areas of uncertainty by surveillance devices. The integer programming model allows for searchers with varying initial positions, search speeds, transit speeds, search endurance, and downtime requirements. To demonstrate its effectiveness, we solve the model to determine an optimal scheme of maneuver for a team of searchers of three different types seeking targets in two areas of uncertainty.

4 - Prioritizing High Schools for Marine Corps Recruiting

Gary Lazzaro, United States Naval Academy, Annapolis, MD, United States

Marine Corps Recruiting Command needs an improved quantitative method to classify high schools for optimizing recruiting efforts. Marine recruiters currently use a qualitative model to classify high schools as priority “1” or “2” or “3” or “N” that is based on school population, number of past recruits, and recruiter opinion. Our analysis shows the current classification process produces inconsistent priority codes for high schools with similar statistics across the nation and even in the same recruiting areas. We use a variety of analytical techniques to assist with the determination of high school priority codes. We use linear regression, machine learning methods, and data wrangling techniques to show how the high school classification process can be improved. Specifically, we demonstrate a data visualization for California high schools overlaid with additional data fields. We merge publicly available United States census tract data for population, income, education levels and veteran population with past Marine Corps Recruiting Command data for each high school into one interactive chart.

SC09

Summit - 329

Social Aspects of Data and AI

Invited Session

Auctions and Market Design

Chair: Azarakhsh Malekian, University of Toronto Joseph L Rotman School of Management, Toronto, ON, Canada

1 - Competing for Limited Attention Online

James Siderius, Tuck School of Business at Dartmouth College, Hanover, NH, United States, Daron Acemoglu, Asuman Ozdaglar

We consider a model of endogenous content creation where articles can consist of two attributes, catchiness and informativeness, one of which is an easily observed characteristic and the other which is hidden and may only be learned through careful reading. A representative platform user with limited attention receives utility from both entertainment and from absorbing information, and strategically chooses how to allocate her time reading based on the headlines she observes. We fully characterize the set of equilibria when M articles can be produced by K creators, which in general will not be unique, admitting a most information rich and least information rich equilibrium. As the number of articles M grows holding K fixed, we find that while the most information rich equilibrium supplies more information, the least information rich equilibrium also supplies less information. As we increase competition (increasing K holding M fixed), information provision deteriorates further across all equilibria because creators cannibalize each other through catchy content to steal user attention, substituting away from more informative content. We show that carefully designed ranking algorithms by the platform can mitigate the quality loss associated with information overload, but that if these algorithms are solely designed to maximize engagement, they tend to exacerbate the problem.

2 - Design of Resale Platforms

Ilan Morgenstern, Stanford University, Palo Alto, CA, United States

We study resale platforms, an emerging type of online marketplaces in developing countries. These platforms are designed for individuals (resellers) to sell products to others as opposed to buying for themselves, enabling them to supplement their income by earning a margin on the transactions they generate. One challenge these platforms face is that competition among resellers may emerge as more of them join the platform, as their social circles increasingly overlap. We investigate the mechanisms by which competition affects resellers and the role of the platform's design in shaping this relationship. We combine analytical modeling with empirical analysis. Our model captures the resellers' interactions with the platform (in searching for products to sell) and with consumers (in offering these products for purchase). We find that as competition intensifies, resellers not only lower their margins but also reduce the effort they exert in searching for products to sell. We empirically validate these equilibrium predictions using data from a major platform: as competition increases, resellers earn lower margins and reduce the number of products they view while browsing on the app, which we interpret as lower search effort. We explore two interventions that aim to benefit resellers: centralizing margin decisions (a practice adopted by several platforms) and optimizing its product ranking algorithm, offering insights to guide key design choices for resale platforms, aiming to foster sustainable online entrepreneurship in developing economies.

3 - Homomorphic Encrypted Revenue Management

Mojtaba Abdolmaleki, University of Michigan, Ann Arbor, MI, United States, Ruslan Momot

We develop a novel homomorphic encryption-based approach to privacy preservation in a dynamic personalized pricing setting. In each period, the firm offers a personalized price to an incoming customer based on (i) this customer's observable characteristics and (ii) the firm's estimate of the demand function (obtained using the data of the historical customers with whom the firm interacted in the past). Our method enables the firm to use homomorphic encryption to encrypt the data of incoming and historical customers, then estimate the demand function and personalize prices directly based on these encrypted data without the need to decrypt them. In contrast to the previous literature, which only preserves the privacy of historical customers via adding statistical noise to their data (so-called statistics-based approach), our approach allows the firm to protect the privacy of all customers -- both incoming and historical. Our theoretical analysis further reveals that our approach i) provides perfect privacy protection (achieving 0-differential privacy) and ii) does so at no cost to the firm's expected revenue, thus achieving better revenue performance than statistics-based algorithms, but (iii) it is computationally expensive. We thus develop a hybrid approach to privacy preservation that leverages the strengths of both statistics- and encryption-based methods, achieving the required privacy protection at a comparatively lower computational cost without significant compromise on the expected revenue. We confirm our theoretical findings through a numerical example based on synthetically generated data.

4 - Measuring Strategization in Recommendation: Users Adapt Their Behavior to Shape Future Content

Jennifer Allen, MIT, Cambridge, MA, United States, Sarah Cen, Andrew Ilyas, Hannah Li, Aleksander Madry

Most modern recommendation algorithms are data-driven: they generate personalized recommendations by observing users' past behaviors. A common assumption in recommendation is that how a user interacts with a piece of content (e.g., whether they choose to "like" it) is a reflection of the content, but not of the algorithm that generated it. Although this assumption is convenient, it fails to capture user strategization: that users may attempt to shape their future recommendations by adapting their behavior to the recommendation algorithm. In this work, we test for user strategization by conducting a lab experiment and survey. To capture strategization, we adopt a model in which strategic users select their engagement behavior based not only on the content, but also on how their behavior affects downstream recommendations. Using a custom music player, we study how users respond to different information about their recommendation algorithm as well as to different incentives about how their actions affect downstream outcomes. We find strong evidence of strategization across outcome metrics, including participants' dwell time and use of "likes." A close analysis of participant behavior rules out experimenter demand as the main driver of these trends. Further, in our post-experiment survey, nearly half of participants self-report strategizing "in the wild," with some stating that they ignore content they actually like to avoid over-recommendation of that content in the future. Together, our findings suggest that user strategization is common and that platforms cannot ignore the effect of their algorithms on user behavior.

5 - When Big Data Enables Behavioral Manipulation

Azarakhsh Malekian, University of Toronto Joseph L Rotman School of Management, Toronto, ON, Canada

We build a model of online behavioral manipulation driven by AI advances. A platform dynami-
cally offers one of n products to a user who slowly learns product quality. User learning depends on a product’s “glossiness,’ which captures attributes that make products appear more attractive than they are. AI tools enable platforms to learn glossiness and engage in behavioral manipulation. We establish that AI benefits consumers when glossiness is short-lived. In contrast, when glossiness is long-lived, behavioral manipulation reduces user welfare. Finally, as the number of products increases, the platform can intensify behavioral manipulation by presenting more low-quality, glossy products.

SC10

Summit - 330

Selected Auctions & Market Design papers from the EC 2024 conference II

Invited Session

Auctions and Market Design

Chair: Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Ali Makhdoumi, Duke University, Durham, NC, United States

Co-Chair: Pengyu Qian, Boston University, Boston, United States

1 - Dynamic Pricing and Learning with Long-term Reference Effects

Shipra Agrawal, Columbia University, New York, New York, NY, United States, Wei Tang

We consider a dynamic pricing problem where customer response to the current price is impacted by the customer price expectation, aka reference price. We study a simple and novel reference price mechanism where reference price is the average of the past prices offered by the seller. As opposed to the more commonly studied exponential smoothing mechanism, in our reference price mechanism the prices offered by seller have a longer term effect on the future customer expectations. We show that under this mechanism, a markdown policy is near-optimal irrespective of the parameters of the model. This matches the common intuition that a seller may be better off by starting with a higher price and then decreasing it, as the customers feel like they are getting bargains on items that are ordinarily more expensive. For linear demand models, we also provide a detailed characterization of the near-optimal markdown policy along with an efficient way of computing it. We then consider a more challenging dynamic pricing and learning problem, where the demand model parameters are apriori unknown, and the seller needs to learn them online from the customers' responses to the offered prices while simultaneously optimizing revenue. The objective is to minimize regret, i.e., the T -round revenue loss compared to a clairvoyant optimal policy. This task essentially amounts to learning a non-stationary optimal policy in a timevariant Markov Decision Process (MDP). For linear demand models, we provide an efficient learning algorithm with an optimal square-root-T regret upper bound.

2 - Incentivized Exploration via Filtered Posterior Sampling

Anand Kalvit, Stanford University, Stanford, CA, United States, Aleksandrs Slivkins, Yonatan Gur

We study "incentivized exploration" (IE) in social learning problems where the principal (a recommendation algorithm) can leverage information asymmetry to incentivize sequentially-arriving agents to take exploratory actions. We identify posterior sampling, an algorithmic approach that is well known in the multi-armed bandits literature, as a general-purpose solution for IE. In particular, we expand the existing scope of IE in several practically-relevant dimensions, from private agent types to informative recommendations to correlated Bayesian priors. We obtain a general analysis of posterior sampling in IE which allows us to subsume these extended settings as corollaries, while also recovering existing results as special cases.

3 - Barriers to Collusion-resistant Transaction Fee Mechanisms

Aviv Yaish, Yale University, New Haven, CT, United States

To allocate transactions to blocks, cryptocurrencies use an auction-like transaction fee mechanism (TFM). A conjecture of Roughgarden [44] asks whether there is a TFM that is incentive compatible for both the users and the miner, and is also resistant to off-chain agreements (OCAs) between these parties, a collusion notion that captures the ability of users and the miner to jointly deviate for profit. The work of Chung and Shi [12] tackles the problem using the different collusion resistance notion of side-channel proofness (SCP), and shows an impossibility given this notion. We show that OCA-proofness and SCP are different, with SCP being strictly stronger. We then fully characterize the intersection of deterministic dominant strategy incentive-compatible (DSIC) and OCA-proof mechanisms, as well as deterministic MMIC and OCA-proof ones, and use this characterization to show that only the trivial mechanism is DSIC, myopic miner incentive-compatible (MMIC) and OCA-proof. We also show that a randomized mechanism can be at most 0.842-efficient in the worst case, and that the impossibility of a non-trivial DSIC, MMIC and OCA-proof extends to a couple of natural classes of randomized mechanisms

4 - Complexity, Communication and Misrepresentation

Junya Zhou, University of Texas at Dallas, Richardson, TX, United States

We investigate how increasing the complexity of the message space, in the presence of limited memory, can reduce misrepresentation in strategic communication. We enrich a standard cheap talk game so that senders must communicate not just a payoff-relevant state, but also payoff-irrelevant attributes correlated with the state. We show that: (i) increasing the set of attributes that may need to be reported (i.e., the complexity of the game) improves the amount of information transmitted in equilibrium, (ii) too much of an increase in complexity leads to a reversal of those gains, (iii) limited memory on the part of players, as well as the relative complexity faced by senders and receivers, drives these changes, and {iv} individuals experience cognitive costs when dealing with complex environments that they are willing to pay to avoid. Our findings demonstrate that the reporting of redundant information may induce equilibria that feature improved outcomes compared to simpler, more direct reporting systems, and point out the importance of complexity when trying to induce truthful information revelation.

SC11

Summit - 331

Network Analysis and Graph Neural Networks I

Invited Session

Telecommunications and Network Analytics

Chair: Wenting Li, Los Alamos National Laboratory, Los Alamos, NM, United States

1 - Complex-Valued Graph Convolutional Neural Networks and Their Applications in Energy Systems

Tong Wu, Cornell University, New York, NY, United States

The effective representation, processing, analysis, and visualization of large-scale structured data over graphs, especially power grids, are gaining a lot of attention. So far most of the literature considered exclusively real-valued graph signals. However, graph signals are often sparse in the Fourier domain, and more informative and compact representations for them can be obtained using the complex envelope of their spectral components, as opposed to the original real-valued signals. This is the case for the AC carrier, and motivates its complex phasor representation. Focusing on applications in power systems, in this paper we generalize graph convolutional neural networks (GCN) to the complex domain, incorporating a complex-valued graph shift operators (GSO) based on the admittance matrix of power grids in the definition of graph filters (GF) and process complex-valued voltage phasor graph signals (GS). The theory developed is generalized to handle spatio-temporal complex network processes. We prove that complex-valued GCNs can be stable with respect to perturbations of the underlying graph support, by bounding of the error propagation through multiple NN layers. The paper showcases the benefits of complex GCN relative to several benchmarks, in power grid state forecasting and cyber-attack detection and localization.

2 - Demystifying and Mitigating Unfairness for Learning over Graphs

Yanning Shen, The University of California, Irvine, Irvine, CA, United States

We live in an era of big data and ``small world'', where a large amount of data resides on inter-connected graphs representing a wide range of physical and social interdependencies, e.g., smart grids and social networks. Hence, machine learning (ML) over graphs has attracted significant attention and has shown promising success in various applications. Despite this success, the large-scale deployment of graph-based ML algorithms in real-world systems relies heavily on how socially responsible they are. While graph-based ML models nicely integrate the nodal data with the connectivity, they also inherit potential unfairness. Using such ML models may therefore result in inevitable unfair results in various decision- and policy-making in the related applications. To this end, this talk will introduce novel fairness-aware graph neural network (GNN) designs to address unfairness issues in learning over graphs. Furthermore, theoretical understandings are provided to explain the potential source of unfairness in GNNs and prove the efficacy of the proposed schemes. Experimental results on real networks are presented to demonstrate that the proposed framework can enhance fairness while providing comparable accuracy to state-of-the-art alternative approaches for node classification and link prediction tasks.

3 - Provable Efficient Graph Neural Network Learning via Joint Edge-Model Sparsification

shuai zhang, NJIT, Newark, NJ, United States

Due to substantial computational challenges associated with training large-scale graph neural networks (GNNs), various sparse learning techniques have been developed to minimize memory and storage demands. These methods include graph sparsification, which involves sampling a subgraph to reduce data aggregation, and model sparsification, which entails pruning the neural network to decrease the number of trainable weights. While these strategies have demonstrated empirical success in reducing training costs while preserving test accuracy, a comprehensive theoretical generalization analysis for sparse learning in GNNs remains underdeveloped. This presentation will concentrate on the theoretical characterization of joint edge-model sparse learning, exploring its impact on sample complexity and convergence rates towards bounded generalization error. We will also provide analytical justifications for the strategies of selectively sampling critical nodes and pruning low-magnitude neurons, showing how these approaches can reduce sample complexity and improve convergence without compromising test accuracy. Finally, we will discuss our latest progress in the generalization analysis of graph transformers, which examines the optimization and generalization of a one-layer Graph Transformer for semi-supervised node classification. This analysis shows that the training mechanism of Graph Transformers, through self-attention and positional encoding, enhances generalization by making the attention map sparse and promoting the core neighborhood during training.

SC12

Summit - 332

Sustainability and Power System Efficiency

Flash Session

Contributed

Chair: Carlos Mateo Samudi Lezcano, Carnegie Mellon University, Pittsburgh, PA, 15213, United States

1 - Equity and Demand Tradeoffs in Level 2 EV Chargers Siting Optimization

Carlos Mateo Samudi Lezcano, Carnegie Mellon University, Pittsburgh, PA, United States

Electric vehicles are a plausible way to decarbonize the US private vehicle fleet. One of the main obstacles to hinder this transition is the availability of public charging infrastructure. This is particularly true for households that do not possess off-street parking, or low income households where owners cannot afford to install their own chargers. In deciding where to place these equipment, there is a tension between serving those locations that are expected to have high demand in the short term, and placing chargers in locations that lack this infrastructure to be able to adopt electric vehicles. Our work provides a framework to investigate this tension, through a bi-objective mixed integer linear program, and a series of metrics we develop to quantify the tradeoffs between different demographic groups in an ex post analysis. Our results indicate that including equity in charger siting and sizing decisions can lead to a more spread out spatial coverage and a higher number of total households with access to chargers. However, this does not come without drawbacks, since our equity objective prioritizes disadvantaged populations. Hence, as we increase our preference for this objective, we leave other demographics relatively worse off. The method was implemented in the cities of Pittsburgh, PA, and Seattle, WA, and can serve as a tool for energy public policy and urban planning inquiries.

2 - An Electric Vehicle Routing Problem applied for a Chain of Swiss Supermarkets

Reinaldo Garcia, University of Brasilia (UnB) - Director ORLaB (www.orlab.com.br), Brasília, Brazil, Bruno Souza, Ian Amaral

The use of electric vehicles by citizens and companies is already a reality in the world. Moreover, the replacement of vehicles powered by fossil fuels has been motivated by the most of the governments aiming to mitigate carbon dioxide emissions. In this work, a route optimization model that includes the electric vehicles distance travel and their load capacity besides the vehicles battery life is implemented to a chain of supermarkets located in Zurich, Switzerland. The implemented model considered crucial points for the solution such as the labor working hours, vehicle load capacity, battery life and demand stores in order to minimize the total distance traveled by the delivery trucks.

3 - Integrated electric vehicle routing with time window and mobile charging problem for logistics delivery system

Senyan Yang, Beijing University of Posts and Telecommunications, Beijing, China, People's Republic of, Ruiyan Zhang

Electric vehicles have gradually emerged as a green transportation mode for urban logistics, considering the increase in global warming and environmental pollution. However, their limited driving range and short battery life limit their widespread deployment. This study proposes an integrated electric vehicle routing with time window and mobile charging problem for logistics delivery system. In this system, mobile charging vehicles, serving as the moving charging stations, can charge the electric vehicles flexibly at customer locations, which can reduce the energy consumption generated by detour visits. The effect of charging on battery degradation is considered to optimize the charging strategies for extending the battery life. A hybrid heuristic algorithm is developed in the adaptive large neighborhood search framework, embedding a dynamic programming algorithm. Mobile charging strategies are designed to optimize the charging scheme based on the synchronization of the electric vehicles and mobile charging vehicles. The numerical experiments are conducted to explore the impact of battery degradation and mobile charging on the electric vehicle routing and charging solutions. The proposed algorithm performs well in solving large-scale problems, demonstrating its efficiency and accuracy.

4 - Coordinated Scheduling and Adaptive Control in a Shared Charging System with On-demand Mobility and Individually Owned Electric Vehicles

Zihao Zhu, Southeast University, Nanjing,China, China, People's Republic of, Xinlian Yu, Xiaofeng Li

Electric vehicles (EVs) have been widely adopted among both individual users and mobility On-demand (MoD) operators, privately due to their environmental and economic benefits. However, the full potential of EVs has not yet been realized, due to the spatial and temporal mismatches between uncertain charging demands and the availability of charging stations. In contrast to existing studies, which focus on charging of MoD fleet or private EVs separately, we aim to optimize the real-time charging scheduling and fleet management for EMoD considering charging admission of private EVs. A bi-level framework is designed to maximize the revenue of shared charging stations and EMoD fleet, combing deep reinforcement learning (DRL), queuing, and combinatory optimization. At the upper level, the number of EVs for charging, serving, and relocating in the fleet are determined via reinforcement learning, incorporating the expected charging revenue from private EVs which are obtained by a queuing model. In particular, the arrival process of private EVs at each charging station is exogenous, while that of fleet EVs is scheduled based on the passenger trip requests and charging accessibility. At the lower level, an adaptive admission strategy of private EVs as well as real-time dispatching for EMoD fleet are optimized based on the upper-level outcomes.

Numerical studies conducted using Manhattan's dataset in New York, USA demonstrate that the proposed framework effectively enhances both the utilization of charging stations and the fulfillment ratio of passenger trip orders.

5 - Improved Parameterization Schemes for WRF Wind Speeds

Siddhant Srivastava, Carleton University, Gatineau, QC, Canada, Kristen Schell

Wind power forecasting plays a critical role in enhancing the reliability of renewable energy sources; however, significant errors remain in current forecasting methods. The most detailed computational representation of local wind speeds is delivered by the Weather Research Forecasting (WRF) model. Despite the widespread use of the WRF model, its performance in diverse terrains remains suboptimal, which is crucial for grid integration and operational planning. Our study proposes a novel WRF model configuration that improves wind speed prediction accuracy across these varied terrains. The refined model significantly reduces wind speed prediction errors, thereby improving the reliability and viability of wind energy as a major power source.

6 - Optimization of Ship-deployed AUVs Synergistic Scheduling for Offshore Wind Turbines Underwater Foundations Inspection

Xu Han, Harbin Engineering University, Harbin, China, People's Republic of, Yuzhen Hu

Guided by the IMO's GHG reduction strategy and the "dual-carbon" goal, offshore wind power has become vital in renewable energy, and more attention has been paid to the regular inspection of offshore wind turbines (OWTs). The Autonomous Underwater Vehicle (AUV) has significantly improved inspection, but the current technology limits it to independently perform long-distance and complex tasks. We propose a ship-deployed AUVs synergistic mode to cover larger areas inspections in a shorter period. A mixed-integer programming model is developed to optimize the vessels' routes and schedule AUVs' drop and pick-up time. An adaptive large neighborhood search heuristic based on constraint programming called ALNSCP is developed for large-scale instances to help update the current solution during the repair operation. The simulation instances-based computational experiments verify the superiority of the synergistic mode and solution method in improving inspection efficiency. Sensitivity analysis further reveals how AUV debugging time and allowed float time affect inspection efficiency and cost. The experimental results can realize the cost reduction and efficiency of OWTs underwater foundations inspection and provide decision support for relevant practitioners to develop safety inspection plans.

7 - Condition Monitoring and Fault Detection of Gearbox Subsystem in Mid-sized Onshore Wind Turbines using SCADA Sensor Data

Jacob Shusko, University of Texas at Austin, Austin, TX, United States

Intelligent operation and maintenance optimization for wind turbines is a crucial area of research, driven by the growing prevalence of wind energy for electricity generation. The gearbox systems within wind turbines are particularly prone to failures within a five-year timeframe due to strenuous operating conditions. Effective monitoring of these critical systems can facilitate predictive maintenance strategies, mitigating the impact of potential total failures which necessitate expensive repairs and prolonged downtime. In this study, we develop condition monitoring models for key subsystems of mid-sized onshore wind turbines, employing both conventional time series modeling techniques and machine learning-based approaches. To demonstrate the efficacy of these models, we conduct a case study utilizing an open-source dataset comprising three years of sensor data from five mid-sized onshore wind turbines, encompassing a range of fault scenarios.

8 - Approximate Maintenance Policies for Large-Scale Offshore Wind Farms

Morteza Soltani, Clemson University, Clemson, SC, United States, Amin Khademi, Jeffrey Kharoufeh

We consider maintenance strategies for a large-scale, offshore wind farm in which wind turbines progressively degrade over time due to normal usage and exposure to a randomly varying environment. The turbines exhibit both economic and stochastic dependence due to shared setup costs and their common environment. The objective is to minimize the expected total discounted setup, replacement and lost power production costs over an infinite horizon. The problem is formulated using a Markov decision process (MDP) model; however, due to the curse of dimensionality, we examine an approximate formulation that is amenable to solution via column generation. Provided are upper and lower bounds to assess the quality of the approximation, as well as conditions under which the optimal policy is obtained via the approximate value function. The sensitivity of the maintenance policy to environment condition, setup cost and farm size is also discussed.

SC13

Summit - 333

Diversity and Organizational Dynamics

Flash Session

Contributed

Chair: Tarun Jain, 54 KLMDC, 380015

1 - Whose pie is bigger? CEO pay, blockholder earnings, and firm acquisitions: Distributive justice approach

Junghoon Kim, Sogang University, Seoul, Korea, Republic of, Jeongil Seo, Gyeonghwan Lee

In this study, we examined the impact of CEO relative pay to blockholder earnings on CEO's pay justice perception and subsequent decisions. Prior corporate governance research has demonstrated that blockholders with ownership exceeding 5% of total shareholdings can compel their CEOs to produce a bigger pie (i.e., higher corporate earnings) while receiving a small slice (i.e., less CEO pay). However, we propose that blockholders' monitoring of CEO pay may exert a negative second-order effect on CEO's strategic decision-making.

Drawing on organization justice theory and resource-based view, we posit that CEO pay results from a zero-sum distribution of corporate earnings between a CEO and shareholders. Thus, when evaluating justice in their pay, CEO will consider blockholder earnings as essential reference points. Because CEOs generally expect to appropriate a larger share of corporate earnings than any other resource provides, they will perceive distributive injustice when their pay is less than blockholder earnings. Relatively underpaid CEOs will attempt to restore pay justice by pursuing acquisitions that have been identified as an effective strategy for augmenting CEO pay. By analyzing a longitudinal sample of S&P 1500 firms spanning from 2001 to 2019, we discovered a positive association between CEO underpayment relative to blockholder earnings and acquisition activities. Moreover, this association is stronger when an underpaid CEO is powerful. Our study significantly contributes to the literatures on CEO pay and organization justice.

2 - Common Ownership Networks and the Expansion of Female Board Representation

Calvin Chiou, National Chengchi University, Taipei City, Taiwan

This study investigates the potential of common ownership to expand the pool of female director candidates and alleviate the underrepresentation of women on corporate boards. Using the California Senate Bill No. 826 and peer female director appointments as quasi-experimental settings, our findings demonstrate that firms under common ownership exhibit greater female board representation and a higher propensity to appoint new female directors. Furthermore, these firms show an increased likelihood of women attaining leadership positions on board committees. Our results remain robust to accounting for the activeness of the Big Three, underscore the network effect of common ownership on the supply side of the labor market for female directors.

3 - ESG Investment Strategy based on ESG Sustainable Growth Using ESG Thematic Revenue Exposure Data

Rei Yamamoto, Keio University, Kohoku-ku, Japan, Akifumi Isogai, Masatoshi Nozaki

This study proposes the use of the degree of diversification of ESG activities as a measure of the sustainability of ESG activities and tests the effectiveness of investment strategies based on it. Specifically, we use ESG-related sales data provided by FactSet to calculate entropy as the degree of diversification of ESG activities. We conduct empirical analysis on the U.S. stock market, and confirm that the proposed strategy yields significant excess returns.

4 - Leave My Case to Harry: The Double Response to Examiner Rejection

Qianqian Liu, Tianjin University, Tianjin, China, People's Republic of, Baofeng Huo

Gendered processes and outcomes are pervasive in the labor market. However, while most studies focus on the impact of applicant gender on their performance evaluations and application outcomes, few of these studies pay attention to the potential impact of evaluator gender characteristics. This leads to our research question: Does the gender of evaluators affect how their decisions are perceived by applicants? If so, how and what factors can moderate this relationship? We examine this research question in the context of patent examination, where applicants can challenge an examiner's final rejection decision by filing an appeal. Our empirical analysis reveals a notable gender disparity: following a rejection-decision sending, female examiners are 6.83% more likely to have their decisions appealed against than their male counterparts. This suggests that applicants may trust the decisions of female examiners less, opting to pass their cases to more advanced reviewers for further examination. Conversely, when male examiners issue the same rejection decisions, applicants prefer to request continued examination with the same examiner rather than filing an appeal. This indicates that applicants are more likely to attribute the rejection to their own application reasons and choose to revise their submissions accordingly when interacting with male evaluators.

5 - Eroding restrictive gender norms through schools: An 8-year follow-up study

Tarun Jain, Indian Institute of Management Ahmedabad, Ahmedabad, India, Diva Dhar, Seema Jayachandran

Abstract

Women and girls continue to lag behind men and boys in many domains, partly due to regressive attitudes about gender roles and the importance of equal opportunities. Evidence suggests that even culturally entrenched gender attitudes are amenable to change, and reshaping attitudes among young people may be a particularly effective strategy to change behaviors and outcomes to become more equitable. We report the impacts of a school-based gender attitude change program in India, evaluated using an RCT methodology, 6 years after the program ended. The evaluation is required since long run effects of the program are uncertain. Gender attitudes could erode over time, or translating the improved attitudes to equitable behavior might be challenging. The hypothesized transmission of attitude change to family members might not occur - because treated individuals might not spend time with their younger siblings discussing attitude change, or because attitudes of adult parents might not be malleable.

In the 6 year follow up, we do not find significant impacts on girls' educational completion, participation in productive activity (enrollment in college or a vocational course, or paid employment), or age of first marriage. The program increased girls' personal autonomy, as female students in treatment schools reported a 0.097 higher standard deviation personal autonomy index than their counterparts in control schools. Younger siblings of treated individuals reported 0.12 standard deviation more progressive gender attitudes overall. Fathers of boys also reported improved gender attitudes, pointing to spillovers from school-based programs to the wider family.

SC14

Summit - 334

Blockchain and Governance

Invited Session

Finance

Chair: Fahad Saleh, University of Florida, Winston-Salem, NC, United States

Co-Chair: Kose John, New York University, New York, NY, United States

1 - Personal Experience Effects across Markets: Evidence from NFT and Cryptocurrency Investing

Chuyi Sun, University of North Carolina, Chapel Hill, NC, United States

I examine how personal experiences causally impact investor behaviors and market boom-bust episodes by exploiting a unique experimental setting in the non-fungible token (NFT) market. Using blockchain transaction-level data for about 1 million wallets, I find that NFT investors who randomly receive more valuable NFTs in the primary market are more likely to participate in subsequent primary market sales and trade more NFTs in the secondary market. These experience effects spill over to the cryptocurrency market as investors who randomly receive more valuable NFTs purchase more lottery-like cryptocurrencies. I also find that personal experiences and new investor inflows have contributed to the formation of bubbles in the NFT market. A model-free reinforcement learning framework best explains the empirical results.

2 - The Economics of Decentralized Autonomous Organizations

Valerie Laturnus, Durham University Business School, Durham, United Kingdom

The advent of blockchain, smart contracts, and Web3 has empowered new concepts for equity partnerships with autonomous operating systems and democratic corporate governance. This paper explores 2,377 of such new partnerships and uses detailed transaction data (from 2017 through 2022) to examine the performance of so-called decentralized autonomous organizations (DAOs) on Ethereum. As a result, I find that DAOs with greater participation rates in voting are associated with superior performance. Small members are a prevalent and important class of investors, while the degree of decentralization in DAOs (ownership concentration) plays only a minor role in firm valuation. Overall, DAOs are an effective organizational structure, when members take an active interest in the venture.

3 - Does Vote Trading Improve Voting Outcome?

Konstantin Sokolov, University of Memphis, Memphis, TN, United States, Nirmol Das, Sailendra Mishra

Decentralized autonomous organization design implies that stakeholders vote to express their individual opinions. Nonetheless, this design is disrupted by vote trading. We take advantage of the blockchain data transparency and explore how vote trading affects voting outcome. Our findings indicate that vote trading facilitates the decision-making by better informed stakeholders. Specifically, informed stakeholders use purchased votes to signal the quality of their contributions to the platform and thereby attract the non-purchased votes of uninformed stakeholders. Vote buying typically attracts 51% more non-purchased votes, and the reputation of vote buying stakeholders improves over time. Therefore, it is unlikely that vote trading leads to overselling the platform contributions. We conduct an experiment to confirm the robustness of our findings. Finally, an event study reveals that a demand shock in the market for votes encourages voting by those stakeholders who used to abstain from voting before the shock. Our findings lend support to theoretical and experimental research showing the benefits of vote trading in the absence of the majority rule.

4 - Centralization vs. Decentralization: First Evidence from the Laboratory

Nir Chemaya, UC Santa Barbara, GOLETA, CA, United States, Gabriele Camera, Gary Charness

We study trading networks where governance of payments flows—or, *validation*—relies on a digital registry. Strategic manipulation of the registry causes validation failure and inefficiency. We contrast two architectures: centralized, where validation authority is concentrated in a single participant (e.g., a Central Bank in a traditional monetary network), and decentralized, with a consensus-based validation mechanism (e.g., blockchain networks such as Bitcoin). Both architectures admit multiple Pareto-ranked equilibria. Pre-play communication, a natural feature of a trading environment, may facilitate coordination on efficient play. In our experimental data, decentralization reduced the incidence of validation failures and boosted trading activity. This governance advantage shows that there is scope for decentralization in innovating monetary and payments systems.

SC15

Summit - 335

Advancements in Choice Modeling and Assortment Optimization

Invited Session

Revenue Management and Pricing

Chair: Sumit Kunnumkal, ISB, India, India

Co-Chair: Omar El Housni, Cornell Tech, New York, NY, United States

1 - Store Network Design for Omnichannel Retailing

Mert Cetin, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Victor Martínez-de-Albéniz, Laura Wagner

We investigate the impact of physical store networks on purchase decisions in omnichannel retailing. Using geolocated consumer-level data, we explore how factors like store proximity, store density, assortment breadth, and inventory availability affect consumer behavior across online, offline, and hybrid channels, where orders are placed in-store but fulfilled online. Our analysis distinguishes between the effects of accessibility, such as the number of nearby stores and distance to the closest store, and service quality, including the variety of products available and stock levels. Our findings indicate that while better physical access to stores generally increases overall sales across all channels, the service quality provided by physical stores has a more nuanced effect: it boosts offline sales but decreases online and hybrid sales. This suggests that consumers who have access to well-stocked and varied physical stores may prefer to complete their purchases in person rather than online. To support retailers in optimizing their strategies, we conduct a counterfactual analysis showing that omnichannel retailers should consider maintaining dense store networks to sustain high sales levels. This contrasts with the recent trend of store closures, indicating that a well-distributed physical presence remains crucial for success in omnichannel retailing.

2 - OMGPT: A Sequence Modeling Framework for Data-driven Operational Decision Making

Hanzhao Wang, Imperial College Business School, London, United Kingdom, Kalyan Talluri, Xiaocheng Li

We build a Generative Pre-trained Transformer (GPT) model from scratch to solve sequential decision-making tasks arising in contexts of operations research and management science which we call OMGPT. We first propose a general sequence modeling framework to cover several operational decision-making tasks as special cases, such as dynamic pricing, inventory management, resource allocation, and queueing control. Under the framework, all these tasks can be viewed as a sequential prediction problem where the goal is to predict the optimal future action given all the historical information. Then we train a transformer-based neural network model (OMGPT) as a natural and powerful architecture for sequential modeling. This marks a paradigm shift compared to the existing methods for these OR/OM tasks in that (i) the OMGPT model can take advantage of the huge amount of pre-trained data; (ii) when tackling these problems, OMGPT does not assume any analytical model structure and enables a direct and rich mapping from the history to the future actions. Either of these two aspects, to the best of our knowledge, is not achieved by any existing method. We establish a Bayesian perspective to theoretically understand the working mechanism of the OMGPT on these tasks, which relates its performance with the pre-training task diversity and the divergence between the testing task and pre-training tasks. Numerically, we observe a surprising performance of the proposed model across all the above tasks.

3 - Sequential Choice Model with Representative Products: Behavior, Modeling and Optimization

Yu Sun, Chinese University of Hong Kong, Hong Kong, Hong Kong, Daniel Zhuoyu Long, Ruxian Wang

Modeling decision behavior among multiple choice options has been an active research area for several decades. Typically, classic discrete choice models consider that customers observe all available products simultaneously and select the one with the highest utility (e.g., the widely used MNL and NL models). In this paper, we propose a novel sequential choice model with representative products to describe customers' purchase behaviors, in which the products are grouped in multiple nests, and each nest has a representative product that is viewed in the first stage and other products that are viewed in the second stage. Customers make their purchase decision sequentially following the product information disclosure process in two stages: in the first stage, customers only observe incomplete information about the product set (i.e., limited information of all representative products) and choose a nest with the most attractive representative; in the second stage, customers unlock full information about all products in the selected nest and purchase the product in this nest with the highest utility, where the no-purchase option may exist in both stages and all nests. We employ the widely used MNL choice model in each stage to formulate the sequential choice process. Under this sequential choice model, we first investigate the problem on the selection of representative products. We then investigate the assortment and nest planning problem and establish the optimality of the adjusted revenue-ordered assortments. Finally, we study the pricing problem and show the optimality of adjusted same-markup pricing policies.

4 - Placement Optimization of Substitutable Products

Omar El Housni, Cornell University, New York, NY, United States, Rajan Udwani

Strategic product placement can have a strong influence on customer purchase behavior in physical stores as well as online platforms. Motivated by this, we consider the problem of optimizing the placement of substitutable products in designated display locations to maximize the expected revenue of the seller. We model the customer behavior as a two-stage process: first, the customer visits a subset of display locations according to a browsing distribution; second, the customer chooses at most one product from the displayed products at those locations according to a choice model. Our goal is to design a general algorithm that can select and place the products optimally for any browsing distribution and choice model, and we call this the Placement problem. We give a randomized algorithm that utilizes an α-approximate algorithm for cardinality constrained assortment optimization and outputs a $\Theta(\alpha)/\log(m)$-approximate solution (in expectation) for Placement with m display locations, i.e., our algorithm outputs a solution with value at least $\Omega(\alpha)/\log(m)$ factor of the optimal and this is tight in the worst case. We also give algorithms with stronger guarantees in some special cases. In particular, we give a deterministic $\Omega(1)/\log(m)$-approximation algorithm for the Markov choice model, and a tight $(1-1/e)$-approximation algorithm for the problem when products have identical prices.

SC16

Summit - 336

[tent] Emerging Topics in Pricing and Revenue Management in Retail Industry

Invited Session

Revenue Management and Pricing

Chair: Mehmet Altug, George Mason University, Fairfax, VA, United States

Co-Chair: Aditya Jain, Baruch College, Zicklin School of Business, New York, NY, United States

1 - Buyout Price Optimization for the Rent-to-Own Business

Metin Cakanyildirim, The University of Texas at Dallas, Richardson, TX, United States, Milad Armaghan, Andrew Frazelle, Divakar Rajamani

We study the multidimensional price optimization problem faced by a rent-to-own (RTO) firm, which rents a product and during the rental offers it for sale repeatedly at a sequence of prices forming the buyout price path. We first employ calculus of variations to obtain optimal buyout prices in closed form for a special case. Next, to overcome the nonconcavity of the profit in the general problem, we formulate an equivalent bilevel optimization over the resource utilization and price path. We then transform the inner pricing problem into an efficient deterministic dynamic program (DP). Standard practice in the RTO business is to steeply decrease buyout prices early in the agreement and gradually later. However, we prove for the special case that the optimal prices should optimally decrease gradually early in the agreement and steeply later. Applying our methodology to jointly optimize buyout prices and inventory, our results reveal that higher inventory levels tend to entail higher optimal prices. Last, in a calibrated case study, we again find that prices should optimally decrease gradually early in the agreement and steeply later, validating our insights from the special case. Moreover, our methodology yields approximately a 22% increase in profit relative to industry prices.

2 - If the Blockchain Could Block: Strategic Blockchain Adoption by Manufacturer as Deterrence to the Selling of Counterfeits by Retailer

Jingjing Weng, Fox School of Business, Temple University, Philadelphia, PA, United States, Abhishek Roy, Subodha Kumar

Counterfeiting is a significant concern for many industries, especially in the online channel. We first investigate an online retailer's motivation to sell counterfeits along with the genuine product. We then explore the strategic adoption of blockchain technology by the manufacturer to combat deceptive counterfeit sold by the retailer. Our results show that the retailer is incentivized to sell counterfeits when the production cost is relatively high. However, a higher production cost does not necessarily imply a higher counterfeit rate. Additionally, we demonstrate that the introduction of deceptive counterfeits can sometimes benefits the manufacturer while hurting the retailer. That is, the sales of deceptive counterfeits may lead to win-win, win-lose, lose-win, and lose-lose outcomes for the manufacturer and the retailer. Moreover, we establish the conditions under which the manufacturer adopts blockchain technology. Furthermore, we identify that blockchain technology can serve as a commitment device and thus benefit the retailer by eliminating the prisoner's dilemma region. We extend the base model by relaxing several assumptions, our analysis reveals that our qualitative findings remain consistent with those derived from the base model. Our research provides novel insights for managers and policymakers in combating counterfeits.

3 - Impact of Price Strategy on Returns: A Price Transparency and Valuation Uncertainty Story

Jane Jiang, The Ohio State University, Columbus, OH, United States, Wedad Elmaghraby, Ozge Sahin

We study how discount strategies affect net sales -- sales after accounting for returns. We evaluate two prevalent discount strategies: bundle discounts, which apply discounted prices exclusively to product bundles, and single-item discounts, where the discount extends to all products. Two primary factors are considered in analyzing the discount strategies' performance in enhancing net sales: the leverage effect of bundles at the return stage, and customers' attentiveness to the discount pricing structure. The latter is particularly relevant under bundle discounts, where customers retaining any items must pay full price, a detail that could be overlooked by those inattentive to pricing, leading to an overestimation of bundle values. We adopt a forward-looking approach to customer behavior by modeling the interconnected purchase and return decisions, incorporating post-purchase valuation changes and potential pricing inattentiveness. Partnering with one of Turkey's largest fashion retailers, we apply structural estimation to ascertain how customers reassess product valuations post-purchase and their attentiveness to pricing details. Furthermore, we simulate generic retail scenarios to discern the optimal discount strategy for retailers. Our findings reveal that factoring in returns shifts the retailer's preference in discount strategies. By tailoring bundles to maximize the leverage effect post-purchase, the retailer's net sales can be enhanced by 15.6%. We also find that customers tend to overlook pricing structures, leading to increased returns. Enhancing customer attentiveness to pricing can decrease the retailer's return rates by 20.9%. Moreover, improving customer attentiveness benefits retailers by enabling them to create more versatile bundle offers, further optimizing their sales strategy.

4 - Sales Versus Subscription Business Models in Retail: Price and Assortment Competition

Mehmet Altug, George Mason University, Fairfax, VA, United States, Aditya Jain, Tolga Aydinliyim, Oben Ceryan

Motivated by retailers' introduction of subscription services, we study consumers' self-selection between "buying from seller firms" versus subscribing to renter firms. Modeling price and assortment competition among sellers and renters, we characterize a solution wherein the renter's large assortment at premium prices and the seller's limited assortment split equilibrium demand.

SC17

Summit - 337

Emerging Techniques in Revenue Management and Logistics Planning

Invited Session

Revenue Management and Pricing

Chair: Zhenzhen Yan, Nanyang Technological University, Singapore, Singapore

1 - Privacy-Preserving Inventory Control

Guangyan Gan, NTU, Singapore, Singapore, Hanzhang Qin, Zhenzhen Yan

Data-driven inventory policies can inadvertently expose sensitive information of retailers in the training data (i.e.demand samples), thus susceptible to privacy risks. In this paper, we focus on the privacy issue in the multistage inventory control problem and design data-driven

inventory policies with differential privacy guarantees which provably prevent such risks. We propose the first differential private inventory control algorithms for the multistage inventory control problem in both the backlog setting and lost sales setting, leveraging offline reinforcement learning algorithms with privacy guarantees. Furthermore, we prove that our algorithm preserves privacy and establish the sample complexity (i.e., the number of demand samples needed) bound of two private inventory policies which match a known lower bound for non-private inventory control models. Theoretical results suggest that collecting more data can offer privacy benefits at no additional cost when the sample size is sufficiently large. Experiments further demonstrate the effectiveness of our proposed privacy-preserving inventory policies in achieving satisfactory performance.

2 - Pricing with Adversarial Inventory Constraints

Jianyu Xu, UC Santa Barbara, Santa Barbara, CA, United States, Yining Wang, Xi Chen, Yu-Xiang Wang

We study a dynamic pricing problem under the online framework, where the potential demand is dependent on the price proposed by the seller (we) at each time period t=1,2, ..., T. However, a perishable inventory is imposed at every time t, *censoring* the potential demand if it exceeds the inventory level. To make it more challenging, the inventory quantity γ_t varies over time and is given by *adversary*. To address this problem, we develop a novel and efficient algorithm that achieves $O(\sqrt{T})$ regret for linear noisy demands. This result matches the information-theoretic lower bound, indicating the optimality of our algorithm. Our findings advance the state-of-the-art in online decision-making problems with censored feedback, offering a theoretically optimal solution that can be broadly applied.

3 - Enhancing Avocado Shelf-life Prediction with Machine Learning and Spectral Analysis

Xiaoyue Zhang, National University of Singapore, Singapore, Singapore, Jingyang Li, Xueming Yuan, Mabel C. Chou

Managing the shelf-life of fresh fruits, like avocados, is crucial to minimize spoilage and waste in supply chains. This study collected daily spectral data from 150 avocados until spoilage and applied machine learning techniques—specifically Quadratic Discriminant Analysis (QDA), XGBoost, and LightGBM—to improve avocado shelf-life prediction. Furthermore, we emphasize the significance of applying Moving Average Smoothing (MAS) Transformation to enhance prediction accuracy. Our results highlight the effectiveness of combining spectral data with advanced data processing and analytics methods in accurately forecasting perishable fruit shelf-life, providing important information for optimizing inventory management strategies.

SC18

Summit - 338

Empirical Research in Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Abhishek Deshmane, Georgia Institute of Technology, Atlanta, GA, United States

1 - Operations and Analytics to Improve Welfare in Artisanal Supply Chains

Somya Singhvi, USC Marshall School of Business, Los Angeles, CA, United States, Divya Singhvi, Xinyu Zhang

This research investigates strategies to enhance welfare and productivity in artisanal supply chains, focusing on two critical directions: supervising geographically distributed weavers and using information transparency to increase productivity. First, because weavers work from geographically remote areas and a limited number of supervisors are expected to visit weavers physically, optimizing supervision operations is a key challenge. Using supply chain data, we provide robust empirical evidence that frequent supervisor visits can play a crucial role in improving artisans' productivity. Our results indicate that a one-day decrease in the average number of days between supervisor visits to remote weavers can increase weaving rates and monthly income by 8.6%-11.3%. We also find that (i) visits to looms with difficult-to-weave rugs and (ii) visits that are consistently scheduled have a more substantial positive impact on weavers' productivity. To capitalize on these insights, we propose an optimization framework for scheduling supervisor visits in the supply chain using a Mixed Integer Linear Program (MILP) and show its impact in practice. Second, we investigate the role of information transparency in improving productivity within artisanal supply chains. By providing weavers with transparent access to information regarding work completion and potential increase in income, we hypothesize that productivity will increase. This transparency could also mitigate severe financial stress that weavers often face during the course of the month. These insights offer practical implications for policymakers and supply chain managers aiming to improve welfare and efficiency in artisanal supply chains.

2 - Choice Models and Permutation Invariance: Demand Estimation in Differentiated Products Markets

Amandeep Singh, University of Washington, Seattle, WA, United States, Ye Liu, Hema YOGANARASIMHAN

Choice modeling is at the core of understanding how changes to the competitive landscape affect consumer choices and reshape market equilibria. In this paper, we propose a fundamental characterization of choice functions that encompasses a wide variety of extant choice models. We demonstrate how non-parametric estimators like neural nets can easily approximate such functionals and overcome the curse of dimensionality that is inherent in the non-parametric estimation of choice functions. We demonstrate through extensive simulations that our proposed functionals can flexibly capture underlying consumer behavior in a completely data-driven fashion and outperform traditional parametric models. As demand settings often exhibit endogenous features, we extend our framework to incorporate estimation under endogenous features. Further, we also describe a formal inference procedure to construct valid confidence intervals on objects of interest like price elasticity. Finally, to assess the practical applicability of our estimator, we utilize a real-world dataset from S. Berry, Levinsohn, and Pakes (1995). Our empirical analysis confirms that the estimator generates realistic and comparable own- and cross-price elasticities that are consistent with the observations reported in the existing literature.

3 - Data-Driven Real-TIME Coupon Allocation in Online Platform: Theory and Experiment

Weiming Zhu, The University of Hong Kong, Hong Kong, Hong Kong, Hanwei Li, Jinglong Dai, Jianfeng Lin, Binqiang Huang

Recent advancements in machine learning and the availability of abundant customer data enable platforms to provide customized coupons to individuals. In this study, we partner with Meituan, a leading shopping platform, to develop a real-time, end-to-end coupon allocation system that is fast and effective in stimulating demand while adhering to marketing budgets when faced with uncertain traffic from a diverse customer base. Leveraging comprehensive customer and product features, we estimate CVR under various coupon values and employ isotonic regression to ensure the monotonicity of predicted CVRs with respect to coupon value. Using calibrated CVR predictions as input, we propose a Lagrangian Dual-based algorithm that efficiently determines optimal coupon values for each arriving customer within 50 milliseconds. We theoretically and numerically investigate the model performance under parameter misspecifications and apply a control loop to adapt to customer features in real-time to better adhere to the marketing budget. Finally, we demonstrate through large-scale field experiments and observational data that our proposed coupon allocation algorithm outperforms traditional approaches in terms of both higher conversion rates and increased revenue. To date, our framework has been implemented by Meituan's bike-sharing division and has distributed coupons to over 100 million users across more than 110 major cities in China. This implementation has resulted in a 0.7% increase in revenue, which is equivalent to an additional 8 million RMB in annual profit.

4 - Unlocking the Power of Exchangeable Tickets: a Study of Consumer Behavior in spORts Events

Ovunc Yilmaz, University of Colorado Boulder, Boulder, CO, United States, Hayri Alper Arslan, Yao Cui

Using transactional data from a US professional sports company that owns a baseball and a soccer team, we investigate the impact of selling exchangeable tickets that provide fans flexibility to change the game they attend for a small additional fee. We find that the use of exchangeable tickets leads to significant changes in customers' purchase time and total spending.

5 - Managing Cart Abandonment: Evidence from an Online Delivery Platform

Can Kucukgul, Rutgers University - Camden, Philadelphia, PA, United States, Gad Allon, Dmitry Mitrofanov

In the era of e-commerce, the phenomenon of cart abandonment has become a significant challenge for online retailers. Reduction in abandonment rates allows retailers not only to recover lost revenue but also to improve customer satisfaction. Utilizing a large-size consumer click data from a prestigious market-leader delivery platform, in this paper, we first aim to uncover both store and cart-level characteristics impacting abandonment. Further, employing a hazard-logit model, we investigate the intricate interplay between cart abandonment and customer shopping behavior. Our first set of results indicate that sticker shocks fail to fully explain why customers are deterring from checking their carts. In addition, we provide valuable insights for online delivery platforms on how to tailor their data-driven promotion strategies to different customer segments.

SC19

Summit - 339

Revenue Management: From Theory to Practice

Invited Session

Revenue Management and Pricing

Chair: Georgia Perakis, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Manuel Moran-Pelaez, MIT Operations Research Center, Cambridge, MA, 02141, United States

1 - Price restraining policies and search cost: economic analysis and implications

Roman Kapuscinski, Ross School of Business, Ann Arbor, MI, United States, Ozge Sahin, Mojtaba Abdolmaleki

During the last twenty years, many traditional retailers have been facing an increasing competition from online retailers and local discounters. Customers are able to experience products in a brick-and-mortar store but purchase online for lower prices. As a result, traditional retailers' sales decrease and some of them stop promoting or carrying such products. For manufacturers, however, traditional retailers play a crucial role by showcasing and advertising products to customers. Manufacturers use price restraining policies to protect retailers' margins. Resale Price Maintenance (RPM) and Minimum Advertised Price (MAP) are two most commonly-used policies intended to protect retailers' margin. Under RPM policy, the manufacturer sets a minimum price for each product and requires all retailers not to price below it. Under MAP policy, manufacturers set a "suggested" retail price. While retailers can sell at lower prices, they are not allowed to advertise a price lower-than-suggested retail price. In this paper, we build a stylized model to study RPM and MAP under various market situations. In particular, we explore which policy is more beneficial for the manufacturer, retailers, and consumers. We find that MAP policy is beneficial for the manufacturer when the search cost is high and consumers are highly heterogeneous in their valuation of the product. Otherwise RPM policy is preferred. Traditional retailers and consumers may prefer MAP to be used, and prefer a market with a higher search cost and higher heterogeneity in consumer valuations, compared to the manufacturer. Online retailers always prefer MAP policy.

2 - A Revenue Management Approach to Fair Hospital Diagnostic Service Scheduling

Joyce Luo, Massachusetts Institute of Technology, Cambridge, MA, United States, Maureen Canellas, Dessislava Pachamanova, Georgia Perakis

The management of diagnostic imaging services is a challenging and important task for hospital systems. Suboptimal scheduling of diagnostic services increases patient wait times and prevents the availability of overnight beds that could otherwise be utilized by incoming patients. Patients are served by a limited number of radiology technicians, who have been experiencing higher rates of exhaustion and burnout. Current diagnostic scheduling systems overlook the importance of distributing scans fairly among technicians, which could help alleviate these problems. In collaboration with a large Level I Trauma hospital in Central Massachusetts, we develop a real-time optimization algorithm that manages daily radiology technician schedules more fairly and efficiently than current practice. Our algorithm balances the need to maximize the number of patients receiving scans, with fairness considerations for technicians in terms of scan load and difficulty. We also consider that patients have different urgency levels and characteristics, so certain more urgent patients need to be seen sooner. We provide analytical guarantees in comparison to a full-knowledge formulation and analyze our algorithm's performance using echocardiogram order data from our hospital partner. Simulation results show that in higher demand scenarios, our algorithm's policy shortens patient wait time and improves throughput for certain patient types compared to current hospital policy. In addition, our algorithm's recommended allocations are more fair for technicians compared to current hospital allocations in terms of total scan time, total number of scans performed, difficult scan time, and number of difficult scans performed.

3 - Bayesian Demand Learning and Revenue Management under Limited Capacity

Mihalis Markakis, IESE Business School, Barcelona, Spain, Victor Martínez de Albéniz, Marcos Serrano

In Revenue Management (RM) problems with limited capacity, the optimal price or quantity decision is mainly driven by the ratio of supply over total demand during the sales season. Under ex ante uncertainty about certain demand statistics, primarily its rate, the seller not only adjusts pricing or capacity allocation to optimize revenues but also to learn the demand. We thus consider a stylized quantity-based RM problem whereby a fixed amount of capacity needs to be sold within a given horizon, into either a high margin, low volume channel, or a low margin, high volume one. The demand rate in either channel is uncertain, but a prior distribution over it is available. We formulate the dynamic optimization problem with Bayesian demand learning. We provide a clean and intuitive structural characterization for the general case of the problem; a closed-form solution for the special case where there is one unit of capacity to sell (precisely the regime of limited capacity); and an efficient heuristic policy for the multi-unit case, which provides near-optimal performance in various regimes in our numerical experiments. We find that higher uncertainty regarding the demand rate may push the seller to opt for a high margin, low volume position for longer, rather than look for higher volumes to accelerate learning, as intuition may suggest. Finally, we show that the monetary value of Bayesian demand learning is comparable to the value of allocating capacity in an optimal way, suggesting that demand learning could be a first-order consideration in RM.

4 - Robust One-Shot Price Experimentation

Ali Daei Naby, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Setareh Farajollahzadeh, Ming Hu

We assess the benefit of one-shot price experimentation for a seller in setting a price with limited information about the customer valuation distribution. Specifically, we consider a seller who knows the exact sales probability associated with a single historical price and aims to maximize the worst-case revenue ratio compared to an oracle with complete knowledge of the valuation distribution, referred to as the competitive ratio. For the class of regular distributions, we assess the benefit of one-shot price experimentation for the seller, who can first choose an experimental price and then set a final price based on the realized sales probability. We quantify the value of experimentation, defined as the difference between the optimal competitive ratio with and without experimentation. More specifically, first, we fully characterize the optimal distributionally robust experimental price point and its guaranteed performance. Second, we quantify the value of experimentation. For example, when the market share of the historical price is 0.05 (resp., 0.75), the performance of the optimal robust price without experimentation is 0.37 (resp., 0.25), while with one-shot price experimentation, the performance is 0.63 (resp., 0.5), where the value of experimentation is 0.26 (resp. 0.25). Third, we provide insights into experimenting with higher or lower prices compared to the historical price. For instance, we show that when the historical sales probability exceeds a threshold at around 0.3, the firm should opt for experimenting with a higher price; otherwise, it should opt for a lower experimental price.

5 - A Robust Optimization Approach to Assortment Planning with Cross-Item Effects

Georgia Perakis, Massachusetts Institute of Technology, Cambridge, MA, United States, Manuel Moran-Pelaez

Assortment planning in retail involves the strategic selection of products to offer customers and their inventory levels, with the goal of maximizing profit. While cannibalization or more generally substitution effects are widely studied in the assortment planning literature, complementarity effects remain relatively unexplored. Cannibalization occurs when a product is bought instead of a similar one, and complementarity occurs when two products are bought together because they will be used together. This paper introduces a novel demand model that captures cannibalization and complementarity effects. We also introduce a fixed point method to compute solutions for the assortment planning problem under the aforementioned demand model. Additionally, we introduce a novel robust method to deal with cross-item effects and their uncertainty while using much less computation power than the fixed point method. Through experiments on synthetic and real-world retail datasets, we demonstrate the effectiveness of our proposed approaches in improving assortment planning outcomes, particularly in scenarios with complex interdependencies between items. Our findings underscore the importance of considering both cannibalization and complementarity effects in assortment planning and provide practical insights for retailers.

SC20

Summit - 340

Predictive Analytics for High-Stakes Decision Making

Invited Session

Decision Analysis Society

Chair: Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Wenbin Zhou, Carnegie Mellon University, Pittsburgh, PA, 15213, United States

1 - Point Processes with Event TIME Uncertainty

Tingnan Gong, Georgia Institute of Technology, Atlanta, GA, United States, Xiuyuan Cheng, Yao Xie

Point processes are popular instruments in statistics that uncover the temporal pattern of dependent event data. In many applications, the event data cannot be observed exactly. Hence, we need to take the time uncertainty into account. Most existing works yet propose point process models with temporally stationary kernels, including the classical Hawkes process. However, modern data demands more general point process models, which can characterize non-stationary temporal influences. To settle the mentioned challenges, we construct a novel point process model that can handle the time uncertainty in data and capture non-stationary temporal influences. We also extend the model to a network setting, which can unveil the temporal causal relationship between nodes. We derive two effective vector fields to learn the kernel and provide theoretical performance guarantees. The statistical performance of the proposed model is demonstrated on simulated and real data, compared with classical and up-to-date baselines.

2 - Sequential Conformal Prediction with Spatial Temporal Point Processes

Wenbin Zhou, Carnegie Mellon University, Pittsburgh, PA, United States

Spatial-temporal forecasting is crucial for various real-world applications, including energy planning, traffic modeling, and epidemic modeling. However, uncertainty quantification methods specifically tailored for spatial-temporal forecasting are limited, primarily due to challenges such as temporal covariate shift and spatial multimodal heterogeneity. In this work, we present a novel conformal prediction framework designed to quantify spatial temporal uncertainty in a distribution-free manner. We assess our proposed framework using both synthetic and real-world datasets, demonstrating that our method achieves a lower coverage gap with high probability and better computational efficiency, compared to state-of-the-art methods.

3 - Towards a Network-Scale Causal Understanding and Forecasting of Traffic Crash Risks with Large Language Models

Frank Yang, Johns Hopkins University, Baltimore, MD, United States, Zhiwen Fan, Pu Wang, Yang Zhao, Yibo Zhao

Current research in traffic accident analysis has predominantly approached the problem as classification tasks, focusing mainly on learning-based classification or ensemble learning methods. These approaches often overlook the intricate relationships among the complex human and contextual factors related to traffic crashes and risky situations. In contrast, we initially formulate a large-scale traffic accident dataset consisting of abundant textual and visual information. We then present a novel application of large language models (LLMs) to enhance causal understanding and improve forecasting of traffic crash specifics, such as accident types and injury severity. Utilizing this rich dataset, we calibrate and fine-tune various LLM configurations to predict detailed accident outcomes based on contextual and environmental inputs. Our AccidentLLM diverges from traditional ensemble machine learning models by leveraging the inherent capabilities of LLMs to parse and learn from complex, unstructured data, thereby enabling a more nuanced analysis of contributing factors. Preliminary results indicate that our LLM-based approach not only predicts the severity of accidents but also classifies different types of accidents and predicts injury outcomes, all with an accuracy above 92%.

4 - Quantifying Uncertainty in Deep Spatiotemporal Forecasting

Dongxia Wu, University of California San Diego, La Jolla, CA, United States, Liyao Gao, Xinyue Xiong, Matteo Chinazzi, Alessandro Vespignani, Yi-An Ma, Rose Yu

Deep learning is gaining increasing popularity for spatiotemporal forecasting. However, prior works have mostly focused on point estimates without quantifying the uncertainty of the predictions. In high stakes domains, being able to generate probabilistic forecasts with confidence intervals is critical to risk assessment and decision making. Hence, a systematic study of uncertainty quantification (UQ) methods for spatiotemporal forecasting is missing in the community. In this paper, we describe two types of spatiotemporal forecasting problems: regular grid-based and graph-based. Then we analyze UQ methods from both the Bayesian and the frequentist point of view, casting in a unified framework via statistical decision theory. Through extensive experiments on real-world road network traffic, epidemics, and air quality forecasting tasks, we reveal the statistical and computational trade-offs for different UQ methods: Bayesian methods are typically more robust in mean prediction, while confidence levels obtained from frequentist methods provide more extensive coverage over data variations. Computationally, quantile regression type methods are cheaper for a single confidence interval but require re-training for different intervals. Sampling based methods generate samples that can form multiple confidence intervals, albeit at a higher computational cost.

SC21

Summit - 341

Business Intelligence and Technology Innovation

Invited Session

Decision Analysis Society

Chair: Nima Safaei, Gies College of Business, University of Illinois Urbana Champaign, Champaign, IL, 61820, United States

1 - Behavioral Analysis and User Motivation in Crowdfunding Campaigns: Effect of HCI driven design

Saunak Basu, University of Nevada Las Vegas, Las Vegas, NV, United States

This study investigates the impact of Human-Computer Interaction (HCI) design on user motivation within crowdfunding campaigns. By analyzing data from diverse crowdfunding platforms, we explore how interface design influences donor engagement and contribution behavior. Our research reveals the significance of HCI-driven elements such as interface usability, persuasive design, and social proof in shaping donation decisions.

Through advanced data analytics and behavioral modeling techniques, we examine nuanced patterns in user interactions and funding outcomes. Additionally, we examine the role of psychological factors like reciprocity and perceived efficacy in driving donor engagement. The synthesis of quantitative analysis and qualitative user feedback provides actionable insights for platform designers and crowdfunding practitioners.

This presentation contributes to the understanding of crowdfunding dynamics by elucidating the complex interplay between HCI design principles and user behavior. By leveraging these insights, practitioners can optimize campaign interfaces to enhance user motivation and maximize fundraising success on digital platforms.

2 - The Impact of Supply Chain Cyberattacks on the Performance of Downstream Firms

Shreya Gokhale, University of Illinois Urbana-Champaign, Champaign, IL, United States, Abhijeet Ghoshal, Gopesh Anand, Yang Gao

Supply chain cyberattack incidents have been increasing in number and severity over the last few years. These cyberattacks occur when malicious actors exploit the buyer-supplier relationships between firms to compromise downstream buyer companies through their suppliers' weaknesses. These attacks have the ability to simultaneously compromise multiple organizations in a supply chain and are difficult to curb since they can be caused by a single security vulnerability of an upstream supplier. Despite the increasing gravity of supply chain cyberattack incidents, there is a scarcity of empirical literature investigating how cyberattacks on an upstream supplier firm impact the performance of the downstream buyer firms associated with the supplier. In this paper, we examine the impact of supply chain cyberattack events on the stock market performance of these downstream buyer firms. Using event study and stacked difference-in-difference methods we analyze supply chain cyberattack incidents that occurred from 2020 to 2023. We find that the publicly known downstream buyer firms associated with such compromised suppliers experience a drop in their stock market performance and the impact lasts for around five trading days after the first public revelation about the event. We further hypothesize that this impact could be moderated by the bargaining power of the buyer firm over the supplier firm and find that the downstream buyer firms having higher bargaining power over their suppliers face a more intense stock market reaction.

3 - Machine Learning, Professional Expertise, and Entry-level Jobs

Yuanyang Liu, University of Tennessee, knoxville, TN, United States, Chuanren Liu, Tingliang Huang

How machine learning (ML) technologies change the job requirements for workers? Using over 51 million job postings of S&P 500 companies from 2011 to 2023, we find that firms utilizing ML also increase the requirements for their workers in prior work experience and decision-making related skills. We provide a causal link with a Shift-Share IV estimation strategy based on the text of occupation description and AI patents. Further, such effects exist for not only knowledge workers (e.g., Managers), but also the lower skill occupations that may not require a college degree (e.g., Retail Salespersons and Customer Service Representatives). Our results indicate that, with the ability to handle straightforward cases and processes, ML technologies may shift the ``mass'' of work for people to handle novel scenarios which necessitate greater domain expertise and decision-making. Consequently, the proliferation of ML could lead to a reduction in entry-level job opportunities. With the ever-growing adoption of ML and related technologies, these consequences require proactive career development strategies for individuals and institutions.

4 - Share-the-Road: Exploring the Relationship between Bicycle-sharing and Ridesharing Platforms

Ayush Sengupta, Alfred University, Alfred, NY, United States

In this paper, we explore the relationship between bicycle-sharing platforms such as Citi Bike and ridesharing platforms such as Lyft. The two groups of platforms are both relatively new business models in the transportation service industry and are driven by innovative online technologies. It is debated in academic research and in industry whether the two types of platforms have a complementary or substitution effect on each other. Some argue that the bicycle-sharing platforms may help alleviate the last-mile barrier in urban transportation and benefit ridesharing platforms by providing additional customers from their untapped market and by improving their financial situation. Others suggest that an adverse impact of bicycle-sharing platforms on ridesharing platforms may arise through a reduction in the ridership of the latter, especially over short trips, due to a substitution effect. We focus on answering this question empirically by examining the impact of the entry of Citi Bike on the ridership of all ridesharing platforms, in the context of New York City. We analyze data on ridesharing-platform ridership, measured by the number of their pick-ups. Through our analyses, we examine how Citi Bike’s entry influences ridesharing’s popularity in a large metropolis and how the impact varies across trips with different travel distances. We also analyze how Citi Bike’s entry influences pooled ridesharing trips and ridesharing trips during rush hours. Our findings offer insights for bicycle-sharing and ridesharing-platform companies to make business decisions and for policymakers to devise policies to improve traffic congestion and reduce vehicular emissions.

5 - Predicting Firm Alliances: Leveraging Relational Pluralism in Innovation with Graph Neural Networks

Junho Yoon, The University of Iowa, Iowa City, IA, United States, Gautam Pant

Interfirm alliances in high-tech industries are a synergistic resource-sharing mechanism that can lead to efficient innovation. As an initial critical step in forming alliances, firms conduct partner search, which involves finding other firms that possess complementary technological resources and capabilities. This study proposes an automated alliance prediction framework that can simplify partner search as well as provide valuable intelligence to third parties such as analysts and investors. Our graph neural network-based framework utilizes a key alliance theory on relational pluralism, which refers to multitudes of relationships that exist between firms that can provide an understanding of how firms interact and collaborate. To operationalize our prediction framework, we compile a rigorous firm-level network data set derived from 8,739 alliances between 11,499 firms in 11 high-tech industries through the period of 1990 - 2018. Our theory-driven predictive models incorporate multiple innovation-based relations—interfirm collaborations, human capital flow, and knowledge spillovers—into graph neural networks (GNNs). Our prediction results show that the plurality of interfirm relations collectively contributes to superior predictive performance across varying evaluation metrics, highlighting the practical utility of our predictive models as a partner search instrument. We also discuss the stakeholders who can benefit from our prediction framework in practice, as well as several contributions to the alliance literature.

SC22

Summit - 342

Strategic Mechanisms in Contemporary Markets

Invited Session

Decision Analysis Society

Chair: Linsheng Zhuang, NUS Institute of Operations Research and Analytics, Singapore, Singapore

1 - Joint Bidding and Information Frictions

Shanglyu Deng, University of Macau, Macao, China, People's Republic of, Qiang Fu, Zenan Wu

In public procurement where some firms must cooperate to finish the project, allowing joint bidding by complementary firms is a natural way to promote competition. This paper compares the performance of the joint bidding approach with that of the VCG auction, which allows complementary firms to submit independent bids without forming consortia. Absent information frictions, joint bidding outperforms the VCG auction, achieving full allocative efficiency at lower procurement prices. However, information frictions can hinder joint bidding due to non-assortative matching of firms or strategic behavior within consortia, making it perform worse than the VCG auction in terms of allocative efficiency and procurement price.

2 - Spillovers on Crowdsourcing Platforms

Sanjiv Erat, University of California-San Diego, La Jolla, CA, United States, Xiaofeng Liu

Crowdsourcing platforms, such as Kaggle, facilitate innovation and problem-solving contests, with diverse deadlines, that run over time and compete for the same set of solvers. The current article offers a tractable model of multiple solvers competing in multiple contests over time, and characterizes the equilibrium solver efforts and the resultant quantity and quality of submitted solutions. The model reveals a negative externality between contests arising from solvers splitting their efforts, and a positive externality arising from the possibility of solver's learning from similar contests. Much more interestingly, the model also predicts a nuanced relationship between when the competing contest ends and the strength of these externalities. Building on these predictions, we conduct an empirical analysis of Kaggle contests. Results confirm a negative correlation between the number of competing contests and both submission quantity and quality, a relationship that is particularly pronounced when competing contests end before focal contests. Analysis of contest similarity shows positive externalities from competing contests to the focal contest, and its interaction with the contest deadlines. Specifically, when the competing contests that ends after the focal contest is more similar, we observe higher submission quantity in the focal contest, while when the competing contest that ends before before the focal contest exhibit greater similarity, the focal contest attracts higher submission quality. Beyond the theoretical contribution, our research provides actionable insights for contest organizers and crowdsourcing platforms, offering the first empirical assessment of how differing deadlines and contest similarity relate to the quality and quantity of solver-submitted solutions.

3 - Sourcing Innovation and Production: the Value Of Commitment

Ersin Korpeoglu, University College London, London, United Kingdom, Xiaoshuai Fan, Cuihong Li, Jun Liang

We consider a buyer firm that sources the design and production of an innovative product from two suppliers. The value of a supplier's design depends on the supplier's effort, while the production cost is the supplier's private information. A supplier can potentially produce the design of another supplier, albeit at a switching cost added to the production cost. Hence, a sourcing mechanism for an innovative product should motivate suppliers' innovation efforts to improve the design value and identify low-cost production solutions. We first investigate the ex-post optimal sourcing mechanism: the *no-commitment* mechanism in which the buyer specifies the rules to select suppliers for design and production after receiving the designs submitted by suppliers. After that, we consider two sourcing mechanisms commonly observed in practice, under which the buyer commits to sourcing the design and the production *jointly* or *separately*. We then compare the no-commitment mechanism with these two mechanisms. We also consider an *enhanced-commitment* mechanism in which the buyer commits to using the supplier who provides the best design for production. Our findings provide valuable managerial insights for buyers to manage suppliers' innovation effort and production costs in sourcing mechanism design.

4 - Aggregating OR Interacting?: the Role of Individual Overconfidence

Wenjie Tang, University of Vienna, Vienna, Austria, Yanwei Jia

We compare two approaches of combining individual judgments about an unknown value into a collective estimate when individuals are overconfident: statistically aggregating individual estimates or asking individuals to interact and reach consensus. We are interested in which approach, aggregating or interacting, is better able to alleviate individual overconfidence and achieve a higher accuracy of the collective estimate.

Consistent with existing findings from laboratory experiments, our model shows that, in the absence of individual overconfidence, the value of group interactions, compared to a statistical aggregation of individual estimates, is always positive. However, we find that this is no longer the case when the individuals are overconfident; in particular, the value of group interactions can become negative when individuals' level of overconfidence is high. Moreover, we show that whereas the accuracy of the aggregating approach always increases in the group size, it is not always optimal to have as many interacting group members as possible when these group members are overconfident. Finally, we provide the optimal solutions for the joint decision of how to form collective estimates, aggregating versus interacting, and how to determine group size.

5 - Orchestrating Organizational Politics: Baron and Ferejohn Meet Tullock

Zenan Wu, Peking University, Beijing, China, People's Republic of, Qiang Fu, Yuxuan Zhu

This paper examines the optimal design of the rules that govern the process of dividing a fixed surplus and recognizing a proposer inside an organization. The process is modeled as a multiplayer sequential bargaining game with costly recognition. The designer is endowed with instruments: first, she sets the voting rule, i.e., the minimum number of favourable votes required to approve a proposal; second, she can manipulate the mechanism of proposer recognition, which is modeled as a contest. That is, she can bias the contest for recognition in favor of certain players by imposing multiplicative weights and additive headstart on their outputs. The design objective accommodates diverse preferences. We show that when the designer can flexibly alter both voting rule and recognition mechanism, the optimum can be achieved by a dictator voting rule and a biased recognition mechanism without headstarts, which reduces to a standard biased rent-seeking model. When the recognition mechanism is kept fixed, the optimum may involve headstarts. When the recognition mechanism is kept fixed, a k-majority-rule with k>1 can be optimal.

SC23

Summit - 343

Theory, Algorithms, and Applications of Multi-objective Optimization

Invited Session

Multi Criteria Decision Making

Chair: Margaret Wiecek, Clemson University, Clemson, SC, United States

Co-Chair: Nathan Adelgren, United States Naval Academy, Annapolis, MD, United States

1 - A Benders' Decomposition Algorithm for Multiobjective Linear Optimization

Benjamin Hamlin, Clemson University, Clemson, SC, United States, Margaret Wiecek

We consider a multiobjective linear program (MOLP) suitable for Benders' decomposition. Applying the weighted-sum method to the MOLP yields a parametric linear program which can be solved with a newly developed parametric Benders' decomposition algorithm. We demonstrate the application of the algorithm to a problem in economics.

2 - Line Decomposition for Multiobjective Programs

Emma Soriano, Clemson University, Clemson, SC, United States, Margaret Wiecek, Mishko Mitkovski

We present a modification of a decomposition theorem for computing the efficient set of multiobjective programs we have earlier established. The decision space, formerly decomposed into subsets, is now decomposed into lines. The efficient sets on these lines can be obtained in a closed form by solving multiple single-objective programs. The new theorem shows that the entire efficient set can be computed as the intersection of the efficient sets on lines. While in general this intersection is infinite, we identify the special cases in which it is finite. Convergence results allowing for a decomposition implementation and examples are included.

3 - On Subproblem Tradeoffs in Multiobjective Optimization

Philip de Castro, Clemson University, Clemson, SC, United States, Margaret Wiecek

We are interested in obtaining efficient solutions for multiobjective optimization problems (MOPs) with many criteria by decomposing an MOP into a set of multiobjective subproblems (MOSPs). Notably, we show that (weakly) efficient solutions of the subproblems can be used to construct (weakly) efficient solutions for the original problem. However, in order to gain a deeper understanding of the interaction between subproblems, we introduce the concept of subproblem tradeoffs which measures the performance of a feasible decision with respect to each MOSP. To compute these tradeoffs, we present extensions to the theory of achievement scalarizing functions. Our new theory leads to an interactive coordination procedure to compute a preferred MOP-efficient solution, and if necessary, achieve an acceptable compromise among the MOSPs. Finally, we demonstrate the utility of our proposed procedure on an assignment problem inspired by humanitarian aid management.

4 - Optimization- and Wargaming-Based Approaches to Improve Medevac Mission Planning and Decision Making

Armin Fugenschuh, Brandenburg University of Technology, Cottbus, Germany

In the highly dynamic environment of military operations, efficient medical evacuation (MEDEVAC) of casualties is a critical strategic challenge. This talk presents a method for planning MEDEVAC missions using mathematical optimization and gamification. By applying Mixed-Integer Programming (MIP), a mathematical optimization technique, to solve a MEDEVAC dispatching problem, a structured approach to decision making in emergency situations is presented. Our MIP model enables the detailed consideration of operational conditions and the mission support with combat helicopters as escorts, and thus improves the strategic planning and efficiency of MEDEVAC operations. We also discuss ethical considerations in prioritizing evacuation decisions and emphasize the need to integrate ethical principles into the development. When it comes to prioritize mission safety versus execution speed, we will enter the realm of multicriteria optimization models. Moreover, we discuss the potential impact of technologies such as wearable monitoring devices on future triage and MEDEVAC planning that could further improve the precision and effectiveness of medical evacuation.

As a further aspect, the role of gamification is emphasized by developing a board game that simulates the planning task and thus promotes understanding of the planning software. This approach allows planners to compare the quality of their manual planning with the computer-generated solution, improve their own planning skills and realize the benefits of software support for automated, AI-assisted planning.

5 - A Multiobjective Exploration of Compact Ecological Reserve System Design

Nathan Adelgren, United States Naval Academy, Annapolis, MD, United States, Lakmali Weerasena

In this work we consider the design of compact ecological reserve systems. Various metrics that have been used in the past to measure the compactness of such a system are analyzed, and new alternatives are proposed. We then discuss modeling techniques that can be used to incorporate a number of the most suitable of these metrics into multiobjective optimization models containing binary variables. We note that many of these metrics require the use of nonlinear objectives and/or constraints. As such, during computational testing we consider both the original forms of these models as well as linearized alternatives. In order to solve these models, we make slight modifications to an existing technique for solving multiobjective integer programs and develop a parallel implementation. Details of the implementation are provided along with computational results.

SC24

Summit - 344

Innovative Approaches in Stochastic and Explanatory Modeling

Contributed Session

Contributed

Chair: Libo Li, University of Southampton, Southampton, United Kingdom

1 - Robust SHAP Computation for Problem-Level Explanatory Modeling

David Collins, Edward Jones, St Louis, MO, United States, Laura Kang

SHAP values are frequently used to understand the order, direction, and relative magnitude of the influence of the underlying features in a model, however trained models of similar quality can have significant differences in SHAP values, which limits the ability to treat them as explanatory of the underlying problem rather than just the model. This paper develops a technique for including SHAP values in the hyperparameter optimization process, which results in a process that is more robust to perturbations in the training data and closer approximations of the true values of an underlying generative process. This algorithm is broadly applicable to a wide range of models, and can be computed efficiently using existing optimization techniques, without a significant decrease in the quality of the model.

2 - Optimal solutions with bounded inequality

John Hooker, Carnegie Mellon University, Pittsburgh, PA, United States, Ozgun Elci, Peter Zhang

When multiple stakeholders are affected by the solution of an optimization model, fairness may be a concern. We investigate the utility distributions across stakeholders that result from maximizing total utility subject to a resource constraint and a bound on inequality, the latter measured by the utility range or the Gini coefficient. We find that in both cases, the optimal solution consists of only two or three utility levels, and if there are three, those on the lowest level receive nothing. We provide closed form solutions that depend on the resource costs of generating utility for individual stakeholders. These results suggest that the occurrence of two or three major socioeconomic classes in many societies may be partially rooted in the mathematics of optimal distributions.

3 - Managing Distributional Ambiguity in Stochastic Optimization through a Statistical Upper Bound Framework

Jian Hu, University of Michigan - Dearborn, Dearborn, MI, United States

Stochastic optimization is often hampered by distributional ambiguity, where critical probability distributions are poorly characterized or unknown. Addressing this challenge, we introduce a new framework that targets the minimization of a statistical upper bound for the expected value of uncertain objectives, facilitating more statistically robust decision-making. Central to our approach is the Average Percentile Upper Bound (APUB), a novel construct that simultaneously delivers a statistically rigorous upper bound for the population mean and a meaningful risk metric for the sample mean. The integration of APUB into stochastic optimization not only fortifies the process against distributional ambiguity but also reinforces key data-driven decision-making attributes, such as reliability, consistency, and comprehensibility. Notably, APUB-enriched optimization problems feature tractability, with particular advantages in two-stage stochastic optimization with random recourse. Empirical demonstrations on two-stage product mix and multi-product newsvendor benchmark problems reveal the benefit of the APUB optimization framework, in comparison with conventional techniques such as sample average approximation and distributionally robust optimization.

4 - A Framework for Stochastic Fairness in Dominant Resource Allocation with Cloud Computing Applications

Jiaqi Lei, Northwestern University, Evanston, IL, United States, Sanjay Mehrotra

Allocation of limited resources often requires fairness considerations. Such situations arise in application areas such as computer systems, health systems, and humanitarian logistics. This paper introduces a fairness framework in multi-resource allocation when the anticipated needs are stochastic. We apply the distributionally robust optimization (DRO) model to the proposed stochastic fairness model when the distribution of the required resources is uncertain. We show that it satisfies key properties of Stochastic Pareto-efficiency, Stochastic sharing incentive, and Stochastic envy-freeness under suitable conditions. Empirically we find that the variance in the requested resources has implications on resource allocation. Convergence and differences in solutions are illustrated using data from two cloud computing applications.

5 - Explainable Artificial Intelligence via generative agent feedback

Libo Li, University of Southampton, Southampton, United Kingdom

Explainable Artificial Intelligence (XAI) empowers modern AI systems by enhancing their ability to provide clear and understandable explanations of system outputs. Interactive machine learning, as a design option for producing XAI, shows promising progress in enriching the machine learning paradigm and offering interfaces and contexts for developing machine learning use cases. Among various research paradigms developing adaptive and incremental learning algorithms, we propose a novel large language model (LLM)-based generative adaptive learning approach. Using LLMs as agents to respond to specific modelling tasks, we evaluate machine learning outputs based on given metrics and embed the evaluation process to refine model building. We explore LLMs' role as generative agents and extensions such as Supervised Fine-Tuning (SFT) and Retrieval-Augmented Generation (RAG) to assess whether extended architectures improve model usability compared to general-purpose LLMs. We also assess the impact of auxiliary data sources on our use cases. The rich information generated during the interactive process benefits system designers and business analysts in gaining insights into model performance. We discuss the implications of our work through case studies in healthcare risk management, product diffusion, and knowledge management. Stakeholders such as system developers, administrators, and end-users stand to benefit from the generative adaptive learning approach in XAI development and deployment.

SC25

Summit - 345

PSOR Best Video Award Finalists

Award Session

Public Sector OR

Chair: Akshaya Suresh, RAND Corporation, Washington, DC, United States

Co-Chair: Scott Rodilitz, UCLA Anderson School of Management, Culver City, CA, United States

1 - Reducing Food Waste in Production: A Field Experiment in Ghana

Mahyar Eftekhar, Arizona State University, Tempe, AZ, United States, Xinming Liu, Karen Zheng, Richard Boso

Research indicates that food waste is a global issue affecting countries across various income levels, with about 931 million tonnes—or 17% of consumer-available food—being discarded in 2019. Meanwhile, nearly 3 billion people worldwide cannot afford a nutritious diet. This study showcases a straightforward, practical method aimed at reducing food waste by focusing on the food preparation processes in restaurants, which contribute to approximately 26% of the total waste.

2 - Technology Enabled Agent-choice and Uptake of Social Assistance Programs

Rakesh Allu, Cornell University, Ithaca, NY, United States, Maya Ganesh, Sarang Deo, Sripad Devalkar

Beneficiaries of social assistance programs with transfers of undifferentiated commodities often have a designated agent to collect their entitlements from. This gives monopoly power to agents over beneficiaries. When coupled with weak government monitoring, agents do not have incentives to adhere to stipulated operating guidelines, leading to reduced uptake by beneficiaries. Governments are attempting to break the monopoly by allowing beneficiaries to choose agents. However, the impact of choice on uptake may be limited by lack of alternate agents in beneficiaries' vicinities, restricted ability of agents to compete with undifferentiated commodities, and collusion among agents. Using a reverse difference-in-differences framework on data from a food security program in two neighboring states in India, we find that providing agent choice results in a 6.6% increase in the quantity of entitlements collected by the beneficiary households. The increase in uptake is about four times higher in regions with high agent density compared with those with low agent density. Nearly all of the increase is attributable to new beneficiaries collecting entitlements from their preassigned agent. This is suggestive of agents improving adherence to operating guidelines in response to choice. We find associative evidence for this response in the number of days agents keep their shops open. Governments executing in-kind transfers of undifferentiated commodities are considering replacement of in-kind transfers with cash. Our results indicate that alternate designs of providing choice even in a limited form can present a viable alternative.

3 - Workload Management and Human-Algorithm Collaboration in Child Welfare Screening

Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States, Zhaohui Jiang, Alan Scheller-Wolf

This video provides an overview of the child welfare system in the U.S. and the reporting process. It highlights the challenges within the current system, particularly caseworker stress and high turnover rates, and discusses a potential solution to effectively reduce stress and manage workloads through improved report screening decisions facilitated by enhanced human-AI collaboration.

4 -

5 - Search and Matching for Adoption from Foster Care

Vincent Slaugh, Cornell University, Ithaca, NY, United States, Ludwig Dierks, Nils Olberg, Sven Seuken, Utku Unver

Coming soon

SC26

Summit - 346

Open Innovation & New Product Development

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Pascale Crama, Singapore Management University, Singapore, Singapore

Co-Chair: Gaoyu Xie, George Washington University, Arlington, VA, United States

1 - Open Innovation for Smes

Gaoyu Xie, George Washington University, Washington, DC, United States, Pascale Crama, Janne Kettunen

We develop a game theoretic model to explore the innovation strategies of small and medium-sized enterprises (SMEs), focusing on two critical factors differentiating SMEs from large firms: risk aversion and cost of capital. These frictions cause launch and effort inefficiencies which reduce the innovativeness of SMEs. Our model studies whether building an open innovation ecosystem and encouraging SMEs to collaborate with institutes of higher learning (IHLs) or large partner firms can improve innovation outcomes. By incorporating IHL into the ecosystem, we find that the SME's innovation efforts increase, though launch inefficiency persists. In comparison, cooperating with partner firm reduces launch inefficiency and decreases innovation efforts. We determine whether the SME chooses to engage in open innovation with the IHL or the partner firm based on the innovation characteristics.

2 - Incentivizing Project Managers Competing for R&D Funds

Niyazi Taneri, Cambridge Judge Business School, Cambridge, United Kingdom, Ozge Tuncel, Nektarios Oraiopoulos, Jochen Schlapp

Theory suggests combining shared and individual incentives can align the incentives of managers competing for R&D funds. We characterize the optimal contract that incentivizes project managers to evaluate their projects and truthfully report evaluation outcomes, which determines the firm's resource allocation decision. Through a series of treatments, we then explore the contract's interaction with different business environments. Our experimental results indicate that the optimal contract alone doesn't sufficiently motivate project managers, suggesting that the business environment, paired with the optimal contract, plays a crucial role in moving towards theoretical outcomes.

3 - Composing New Product Development Portfolios with Internal and External Projects

Jochen Schlapp, Frankfurt School of Finance & Management gGmbH, Frankfurt Am Main, Germany, Hossein Nikpayam, Moritz Fleischmann

The process of building portfolios with competing internal and external new product development projects comprises two daunting challenges: (i) the collection of relevant information about the projects under consideration and (ii) the selection of the most promising projects under uncertainty. In order to create optimal NPD portfolios, firms must implement project selection policies that carefully control the tensions that arise between these two challenges. Most notably, installing a higher level of selection flexibility allows firms to better select projects based on the information they obtain; however, greater selection flexibility also induces fiercer competition for resources among the projects, which undermines the reliability of a firm's information acquisition efforts and thus harms the quality of its selection decisions. Hence, we ask the following question: Which project selection policies can best align a firm's (ex ante) information acquisition efforts with its (ex post) project selection decisions when the firm contemplates investing in competing internal and external NPD projects? We study this question by developing a principal-multiagent model, and we identify different contingencies
that strongly impact a firm's optimal project selection policy, including the projects' market potential and riskiness, the quality of acquired information, and the severity of agency issues.

4 - Three Years, Two Papers, One Course Off: Optimal Non-Monetary Reward Policies

Shivam Gupta, University of Nebraska Lincoln, Lincoln, NE, United States, Wei Chen, Milind Dawande, Ganesh Janakiraman

Motivated by practical examples from academia (a reduced teaching-load for achieving a certain research productivity threshold) and industry (Supplier-of-the-Year awards in recognition of excellent performance), we consider a principal whose goal is to design a policy for giving non-monetary rewards to an agent. The following limited-term reward policy has been quite popular in practice: The principal evaluates each agent periodically; if an agent's performance over a certain (limited) number of periods in the immediate past exceeds a pre-defined threshold, then the principal rewards him for a certain (limited) number of periods in the immediate future. We identify a limited-term policy that is arbitrarily near-optimal. We also introduce and analyze the class of score-based policies and show that it always contains an optimal policy.

SC27

Summit - 347

Emerging Topics in NPD and Innovation

Invited Session

New Product Development

Chair: Yi Xu, University of Maryland, College Park, MD, United States

1 - R&D Project Portfolio Collaboration: How to Structure the Strategic Alliance?

Guiyun Feng, Singapore Management University, singapore, Singapore, Pascale Crama, Wenqi Lian

Incumbent companies target innovative firms with promising research and development (R&D) projects to rejuvenate their product portfolio. Such strategic alliances create value by combining the innovator's research expertise with the partner firm's superior marketing capability. The partner firm chooses the timing and payment terms of the strategic alliance, accounting for the innovator's budget and marketing capability, and the project portfolio's market interaction and revenue variability. The partner firm may prefer delayed alliances post R&D completion for innovator firms with high marketing capability when the innovator has sufficient budget or when market interaction is weak. Upfront alliances—prior to the R&D stage—are always preferred when the innovator's marketing capability is low to cement the partner's commitment to payments that incentivize innovator R&D. The partner also contracts upfront when projects exhibit strong market interaction yet the innovator is budget-constrained, as the upfront payment augments the innovator's R&D budget. Finally, a strategic alliance may fail to form when the project portfolio has high revenue variability and low market interaction, and the innovator has low marketing capability. Interestingly, the partner's profit does not always decrease in the innovator's marketing capability—her outside option—and the partner may prefer an innovator with intermediate marketing capability when revenue variability is high. Furthermore, the partner's profit weakly increases in the innovator's R&D budget up to a threshold, yet may exhibit a discontinuous jump/drop at that threshold.

2 - Consumer Privacy and Personalized Recommendation on Online Platforms

Farzad Fathi, University of Maryland, College Park, MD, United States, Yi Xu, Bo Zhou

This paper examines the privacy-personalization paradox within online platforms, focusing on the trade-offs between consumer privacy and personalized recommendations and pricing. Our research incorporates a model that characterizes consumers by their maximum willingness to pay and their product preferences, treating privacy levels as a crucial factor influencing platform decisions on product offerings and pricing strategies.

We emphasize privacy's dual impact on consumer utility and platform profits. The study identifies conditions under which consumers might surrender their privacy to maximize expected utility, particularly when the benefits of increased product consumption outweigh the costs of privacy loss. Through this comprehensive analysis, the paper sheds light on the complex dynamics between privacy, personalization, and consumer behavior in the digital marketplace.

3 - Designing Data Science Contests: the Role of Training vs. Testing Split

Ping-Chieh Huang, University of California, San Diego, La Jolla, CA, United States, Zhe Zhang, Sanjiv Erat

Companies organize data science contests to source innovative solutions that provide long-term competitive advantage. Using contests hosted on Kaggle as a motivating example, we formulate a model of data science contests where the final evaluation is performed with a statistical holdout sample. Our theoretical results illustrate that the split ratio of training vs. testing samples influences the contestants' incentives, and their likelihood of submitting a statistically efficient solution. Specifically, our investigation of 30 most recent contests shows contestants' tendency of pursuing a risk-seeking approach and a potential agency problem between innovation seekers and problem solvers. Our prescriptive analysis of the optimal split ratio manages these agency issues and provides actionable insights for firms hosting data science contests.

4 - Product Line Design and Planning for Startups

Sreekumar Bhaskaran, SMU-Cox School of Business, Dallas, TX, United States, Karthik Ramachandran, Zeya Wang

How should a startup design its product line? Startups face several unique challenges including limited resources, market uncertainty, and diverse customer preferences. Through a model-based investigation, we explore strategies for optimizing market coverage while mitigating the risks of cannibalization in the pursuit of startup success.

SC28

Summit - 348

Empirical Research in Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Tim Kraft, NC State - Poole College of Management, Raleigh, NC, United States

Co-Chair: Junhao Yu, Miami University, Oxford, OH, United States

1 - Is Climate (Adversity) the Mother of Innovation? Evaluating the Impact of U.S. Billion-Dollar Weather and Climate Disasters on Firm-Level Innovation

Max Ji, The Pennsylvania State University, University Park, PA, United States, Suvrat Dhanorkar, Suresh Muthulingam

The escalating frequency and intensity of climate-related extreme events, such as hurricanes, floods, droughts, tornadoes, and wildfires, pose substantial global challenges. The United States, in particular, has experienced an increase in major weather-related disasters and billion-dollar weather-related damages over the past five decades. This underscores the pressing need to understand the impact of these events. This paper focuses on exploring how extreme weather events affect firm-level innovation. We use data from the National Oceanic and Atmospheric Administration (NOAA) and the US Patent and Trademark Office (USPTO) to study the relationship between extreme weather events and innovation within impacted firms. Our results indicate that while climate events generally impede overall innovation, they also stimulate the advancement of green innovation within affected companies.

2 - Pride or Guilt? Impacts of Consumers' Socially Influenced Recycling Behaviors on Closed-Loop Supply Chains

wenlin chen, University of Electronic Science and Technology of China, Chengdu, China, People's Republic of, Wenjie Huang, Jason Nguyen, Chung-Li Tseng, Sam Kirshner

Social influenced pride and guilt are found to impact recycling behavior, but are overlooked in closed-loop supply chain research. In contrast, Environmental Psychology literature ignores firms' operational decisions. We analyze how social influence affects consumers' recycling behaviors and their resulting impacts on a manufacturer's strategy in a closed-loop supply chain.

3 - Sustainable and Affordable? The Effects of CSR Performance and Price on Consumer Purchase Decisions

Tim Kraft, NC State - Poole College of Management, Raleigh, NC, United States, Junhao Yu, Robert Handfield, Rejaul Hasan, Marguerite Moore

We use controlled experiments in an online purchase context to examine how consumers' willingness to buy is influenced by a retailer's disclosure of a manufacturer's CSR performance. We show that disclosing CSR performance is more challenging (and potentially riskier) with consumers who typically pay a higher price, but consumers who typically pay a lower price are willing to reward positive CSR performance.

SC29

Summit - 420

Optimization Society's Award Session II

Award Session

Optimization

Chair: Oktay Gunluk, Georgia Tech, Atlanta, GA, United States

1 - Uniqueness of DRS as the 2 Operator Resolvent-splitting and Impossibility of 3 Operator Resolvent-splitting

Ernest Ryu, UCLA, Los Angeles, CA, United States

Given the success of Douglas–Rachford splitting (DRS), it is natural to ask whether DRS can be generalized. Are there other 2 operator resolvent-splittings sharing the favorable properties of DRS? Can DRS be generalized to 3 operators? This talk presents the answers: no and no. In a certain sense, DRS is the unique 2 operator resolvent-splitting, and generalizing DRS to 3 operators is impossible without lifting, where lifting roughly corresponds to enlarging the problem size. Finally, we conclude by overviewing the recent line of research in optimization focused on characterizing and classifying first-order optimization algorithms.

2 - On the Column Number and Forbidden Submatrices for Delta-modular Matrices

Luze Xu, University of California, Davis, Davis, CA, United States, Joseph Paat, Ingo Stallknecht, Zach Walsh

Totally unimodular (TU) matrices have been used for decades to solve various integer programs (IPs) in polynomial time using linear programming. Given the importance of TU matrices, one may ask about generalizations. One such generalization is the family of Delta-modular matrices, whose full rank subdeterminants are bounded by Delta. An open conjecture is whether IPs with Delta-modular constraint matrices can be solved in polynomial time if Delta is fixed. A natural approach to tackle this conjecture is to understand structural properties of Delta-modular matrices. In this talk, we explore one such structural property: how many distinct columns can a rank-r Delta-modular matrix have? We give an overview of this question and discuss some implications for Delta-modular IPs. By identifying excluded submatrices and combining these with results from matroid theory, we derive a new bound on this column number that is best possible up to a polynomial in Delta.

3 - The Magic of Monotone Integer Programs: Polynomial Time Solvability with Min-cut Procedure and Fast High Quality Solutions for Related Hard Problems

Dorit Hochbaum, UC Berkeley, Berkeley, CA, United States

Monotone Integer problems (IPM) are those formulated as integer programming with constraints that have at most two variables, x-variables, appearing with opposite sign coefficients, and a third variable, z-variable, that can appear in one constraint only. All IPMs are solved as a minimum s,t-cut on a respective graph.

Any NP-hard minimization problem formulated with two variables per inequality, has a 2-approximation algorithm, attained with one min-cut procedure, resulting from
a transformation to IPM.

Many NP-hard problems are IPM plus a budget constraint, e.g. the quadratic knapsack problem and facility dispersion problem. For these, the breakpoints algorithm that relaxes the budget constraint, delivers high quality solutions in very fast run times. This presentation focuses on recent results for IPM ratio problems, that include the maximum density subgraph problem. We introduce the Incremental Parametric Cut algorithm, that solves monotone ratio problems in the complexity of a single min-cut procedure. This algorithm is efficient not only in theory but in practice as well. For large scale instances of the maximum density problems, it is orders of magnitude faster than recent leading methods that are all based on heuristics, and also significantly faster than the theoretically-efficient parametric cut algorithm.

SC30

Summit - 421

Large-Scale Methods for Linear and Nonlinear Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Zikai Xiong, MIT OR Center, Cambridge, 02139, United States

Co-Chair: Robert Freund, MIT Sloan School of Management, Cambridge, MA, 02139, United States

1 - The radius of statistical efficiency

Mateo Diaz, Johns Hopkins University, Baltimore, MD, United States, Joshua Cutler, Dmitriy Drusvyatskiy

Classical results in asymptotic statistics show that the Fisher information matrix controls the difficulty of estimating a statistical model from observed data. In this work, we introduce a companion measure of robustness of an estimation problem: the radius of statistical efficiency (RSE) is the size of the smallest perturbation to the problem data that renders the Fisher information matrix singular. We compute the RSE up to numerical constants for a variety of test bed problems, including principal component analysis, generalized linear models, phase retrieval, bilinear sensing, and matrix completion. In all cases, the RSE quantifies the compatibility between the covariance of the population data and the latent model parameter. Interestingly, we observe a precise reciprocal relationship between the RSE and the intrinsic complexity/sensitivity of the problem instance, paralleling the classical Eckart–Young theorem in numerical analysis.

2 - Level-Set Geometry and Improving the Performance of PDHG for Linear Optimization

Zikai Xiong, MIT OR Center, Cambridge, MA, United States, Robert Freund

It is now the case that many linear programming (LP) problem instances are at a scale where matrix-factorization-free methods are attractive or necessary. The restarted primal-dual hybrid gradient method (rPDHG)—with heuristic enhancements and GPU implementation—has been very successful in solving these huge-scale LP problems; however, its performance can be highly variable and lacks intuitive understanding. We analyze the theoretical foundation and practical performance of rPDHG and generalize it to general conic linear optimization. Our analysis reveals that the geometry of the primal-dual (sub-)level sets plays a crucial role in the performance of rPDHG. Specifically, unfavorable geometry of some instances leads to the poor performance of rPDHG, both in theory and practice. To address this issue, we show how central-path-based linear transformations - including conic rescaling - can markedly enhance the convergence of rPDHG. Furthermore, we present computational results that demonstrate how such rescalings accelerate convergence to high-accuracy solutions, and lead to more efficient methods for huge-scale linear optimization problems.

3 - Optimization on a Finer Scale: Bounded Local Subgradient Variation Perspective

Jelena Diakonikolas, University of Wisconsin-Madison, Madison, WI, United States, Cristobal Guzman

Nonsmooth optimization problems are ubiquitous in industrial and machine learning applications, arising from the need to address tasks such as resource allocation, threat detection, and model training. Within the realm of mathematical optimization, nonsmooth problems stand as some of the most challenging; yet they often offer the possibility of developing efficient algorithms with provable guarantees. The complexity of these problems, encompassing both lower and upper bounds, has historically typically been examined under generic assumptions, bounding the growth of the objective functions by assuming Lipschitz continuity of either the objective itself or its gradient. In this talk, I will argue that these traditional assumptions defining classes of nonsmooth optimization problems inadequately capture local properties of problems that may make them amenable to more efficient algorithms. I will introduce a notion of bounded local variation of the (sub)gradient and discuss how, under this notion, we can obtain a more fine-grained characterization of the complexity of nonsmooth problems. One consequence of these results is that, contrary to prior belief, the complexity of nonsmooth optimization problems, such as those with piecewise linear objectives with polynomially many pieces, can be improved using parallelization even in high dimensions, either if the problem is unconstrained or if the function is allowed to be queried outside the feasible set.

4 - Solving Saddle Point Formulations of Linear Programs with Frank-Wolfe

Matthew Hough, University of Waterloo, Waterloo, ON, Canada, Stephen Vavasis

We discuss the author's recent efforts toward applying the Frank-Wolfe algorithm to solving linear programs. The talk will introduce two first-order primal-dual algorithms for solving saddle point formulations of linear programs, namely FWLP and FWLP-P. The former iteratively applies the Frank-Wolfe algorithm to both the primal and dual of the saddle point formulation of a standard-form LP. The latter is a modification of FWLP in which regularizing perturbations are used in computing the iterates. We will outline our convergence analysis of FWLP-P, noting that FWLP convergence guarantees have not yet been established, and finally describe the advantages of using FWLP and FWLP-P for solving very large LPs.

SC31

Summit - 422

Methods for Large-Scale Nonlinear and Stochastic Optimization I

Invited Session

OPT: Nonlinear Optimization

Chair: Jiahao Shi, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Shagun Gupta, UT Austin, Austin

Co-Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Retrospective Approximation for Stochastic Constrained Problems Using Sequential Quadratic Programming

Shagun Gupta, UT Austin, Austin, TX, United States, Raghu Bollapragada, Albert Berahas

Sequential Quadratic Programming (SQP) is one of the state of the art algorithms to solve deterministic constrained problems. In recent years the framework has been extended to solve equality constrained problems with stochastic objective functions. To adapt the step size in stochastic settings, new schemes like stochastic line search, Lipschitz constant estimation, hessian averaging have been introduced in SQP. We use SQP algorithms in a Retrospective approximation framework that allows us to solve a series of subsampled deterministic subproblems to solve the the stochastic constrained problem. This framework decouples the stochasticity from the SQP algorithm allowing us to use legacy deterministic solvers (other than SQP as well) to solve constrained stochastic programs.

2 - High Efficiency in Stochastic Trust Regions

Sara Shashaani, North Carolina State University, Raleigh, NC, United States, Yunsoo Ha

Trust Region (TR) methods are celebrated for their flexibility in finding stationary points of non-convex problems and their automated step size adjustments. In noisy settings, stochastic and deterministic sources of error are linked via the TR radius, governing the samples needed to estimate the objective function value at each point. But this comes at the cost of significant added complexity. Recent developments such as variance reduction mechanisms and regularization steps have shown promise for making TR competitive with typical worst case complexity lower bounds of non-convex stochastic optimization algorithms. We discuss the features and risks within a TR framework that can guarantee better complexity rates in this literature.

3 - A Double Stepsize Stochastic SQP Method with Complexity Guarantees

Michael ONeill, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Stochastic gradient algorithms play a key role in the training of large-scale machine learning models. These algorithms can be readily extended to problems with simple constraints, such as when projections onto the feasible region can be computed efficiently. Recently, there has been a surge of interest in stochastic gradient-like algorithms for stochastic optimization problems with nonlinear and nonconvex constraints due to modern machine learning applications such as fair machine learning and physics informed neural networks. In this talk, we present a new stochastic Sequential Quadratic Programming (SQP) for deterministically constrained stochastic optimization. Unlike previous work, this algorithm utilizes a different step size for each component of the orthogonal decomposition of the SQP step, which enables faster convergence with respect to the constraint violation and an improved worst-case complexity result. In addition to improving the worst-case complexity, this algorithmic approach also enables significant flexibility with respect to computing step sizes, such as employing a (safeguarded) line search on the constraint violation. Some preliminary numerical experiments will also be presented.

4 - A Quasi-Newton Method for Nonsmooth, Nonconvex, Stochastic Optimization

We consider the minimization of a Lipschitz continuous and expectation-valued function over a closed and convex set. Our focus lies on obtaining both asymptotics as well as rate and complexity guarantees for computing an approximate stationary point (in a Clarke sense) via zeroth-order schemes. We adopt a randomized-smoothing-based approach reliant on minimizing a smooth approximation of our objective function. In such a setting, we develop a zeroth-order stochastic quasi-Newton scheme reliant on a combination of randomized and Moreau smoothing is proposed and analyzed, for which iteration and sample complexities are derived.

SC32

Summit - 423

Recent Advances in Mixed-Integer Nonlinear Programming

Invited Session

OPT: Global Optimization

Chair: Linchuan Wei, Northwestern University, Evanston, IL, 60201, United States

1 - An Outer Approximation Method for Solving Mixed-Integer Convex Quadratic Programs with Indicators

Linchuan Wei, Northwestern University, Evanston, IL, United States, Simge Kucukyavuz

Mixed-integer convex quadratic programs with indicator variables (MIQP) encompass a wide range of applications, from statistical learning to energy, finance, and logistics. The outer approximation (OA) algorithm has been proven efficient in solving MIQP, and the key to the success of an OA algorithm is the strength of the cutting planes employed. In this paper, we propose a new technique for deriving cutting planes for MIQP from various convex relaxations, and, as a result, we develop new OA algorithms for solving MIQP at scale. The contributions of our work are two-fold: (1) we bridge the work on the convexification of MIQP and the algorithm design to solve large-scale problems, and (2) we demonstrate through a computational study on the sparse portfolio selection problem that our algorithms give rise to significant speedups compared with the state-of-the-art methods in the literature.

2 - Mixed Integer Approach for Fair Regression

Anna Deza, University of California-Berkeley, Berkeley, CA, United States, Alper Atamturk, Andres Gomez

TBD

3 - The Treewidth-Convex Hull Theorem and DP for Cut Generation in MINLP

Sourabh Kumar Choudhary, Georgia Tech, Atlanta, GA, United States, Santanu Dey, Nick Sahinidis

We revisit unconstrained nonlinear binary programming problems with a limited treewidth of the associated hypergraph. We provide an alternate proof of exactness of a reformulated LP with $O(d^2n)$ variables, n being the number of variables and d being the treewidth. Our proof links the solution of this exact LP with the well-known dynamic programming algorithm which is known to run in $O(d^2n)$ time. In the second part, we introduce column generation method for solving CGLP to add cuts to non-convex problems with multilinear intermediates. We use the DP algorithm to solve pricing problem. The benefit over the currently practiced algorithm employing graycodes for navigating and pricing columns is experimentally shown.

4 - Disjunctive Sum of Squares

Yixuan Hua, Princeton University, Princeton, NJ, United States, Amir Ali Ahmadi, Sanjeeb Dash, Bartolomeo Stellato

We present the disjunctive sum of squares (DiSOS) approach for certifying nonnegativity of polynomials. Our method strategically partitions the certification problem into tractable subproblems, providing certificates with low polynomial degrees. We introduce the concept of DiSOS cone, and show that it can approximate the set of positive definite forms arbitrarily well. Using the DiSOS cone, we construct a converging hierarchy of optimization problems to solve general polynomial optimization problems. Furthermore, we provide a constructive approach to iteratively partition the certification problem in order to develop efficient solution algorithms. Numerical experiments show that our method is very versatile, as it is able to certify nonnegativity of non-sum-of-squares polynomials, prove copositivity of matrices, and compute the clique number of graphs.

5 - What's new for the global solver in FICO Xpress

Tristan Gally, FICO, Birmingham, United Kingdom

This talk discusses the latest updates and improvements to the global optimization capability within FICO Xpress Solver, which allows to solve general mixed-integer nonlinear problems to global optimality. We will discuss the internal workings of the solver, its features, and recent performance improvements.

SC33

Summit - 424

Journal of the Operational Research Society 75th Anniversary

Invited Session

OPT: Optimization Under Uncertainty

Chair: Haitao Li, University of Missouri - St. Louis, Saint Louis, MO, United States

1 - Multi-Facility Location Models Incorporating Multipurpose Shopping Trips

Pawel Kalczynski, California State University-Fullerton, Fullerton, CA, United States, Zvi Drezner, Morton O'Kelly

This work continues to develop and explore the impact of multipurpose trips on retail location. We develop the model of locating multiple competing facilities of a chain where several competing facilities exist in the area. There may be some existing facilities of the same chain as well. The addition of multiple new outlets can cause cannibalization of existing sales, but this effect is mitigated by selecting good locations, and the total market share captured by the chain increases. The introduction of multipurpose trips enhances the total market share of the location decision maker. Since in reality many customers combine a visit to more than one facility in one shopping trip, the model predicts the expected market share captured more accurately. Therefore, the selected locations for new facilities are more accurate as well.

2 - Efficient Drone Integration for Last-Mile Rural Healthcare Delivery

Shakiba Enayati, University of Missouri - Saint Louis, St Louis, MO, United States, Sina Ansari, Ziteng Wang, Srikanth Gururajan

Rural areas encounter challenges such as limited infrastructure, vast distances, rugged terrains, and harsh weather conditions. These factors impede the timely and efficient delivery of medical supplies, putting the health and well-being of residents at risk. Existing literature recognizes the potential of drones in healthcare logistics, yet a critical research gap exists in optimizing drone selections in the health supply networks considering realistic energy consumption. We develop a decision-support tool to pragmatically integrate drones into existing last-mile distribution networks, considering uncertain environmental conditions such as wind, air quality, precipitation, and terrain impacting drone flight range and payload requirements. We apply our model to a case study from Shield Illinois, an organization that enabled the rapid and widespread distribution of medical supplies, including test kits and samples, during the COVID-19 pandemic, and present our findings.

3 - On the Optimal Flexibility of Stochastic Service Systems with Multi-Class Customers

Zhe Zhang, Simon Fraser University/Western Washington University, Bellingham, WA 98229, USA, WA, United States

In this talk, by developing some stylized queueing models, we will discuss the issue of determining the optimal level of the service flexibility for a stochastic service system with multiple class of customers. Both theoretical results and numerical illustrations are presented to the insights for practitioners for these service systems.

SC34

Summit - 425

Advances in Nonlinear, Stochastic, and Constrained Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Baoyu Zhou, Arizona State University, Tempe, AZ, United States

1 - Universal First-Order Methods for Convex Optimization Problems

Jiaming Liang, University of Rochester, Rochester, NY, United States, Vincent Guigues, Renato Monteiro

We present a generic algorithmic framework for convex optimization without knowing the levels of convexity and smoothness of the problem. In particular, the framework includes an adaptive subgradient method and an adaptive proximal bundle method. One advantage of the framework is that it does not perform any line search on the strong convexity parameter. To the best of our knowledge, this is the first universal method that does not use a backtracking procedure on the strong convexity parameter. We present complexity results in terms of both the optimality gap and the stationarity. Their complexity bounds are as good as those obtained with known convexity and smoothness parameters.

2 - Stochastic Nonlinear Constrained Optimization

Qi Wang, Lehigh University, Bethlehem, PA, United States, Frank E. Curtis

This presentation covers two recently proposed classes of stochastic algorithms for constrained optimization that I have explored in my research: stochastic interior-point methods (sIPM) and stochastic sequential quadratic programming (sSQP). These methods are designed to solve problems where information about the objective can only be accessed through stochastic gradient estimates wheras constraint function and derivative values can be computed explicitly. I will discuss the challenges for ensuring the convergence of these algorithms and how our methods overcome these challenges. Applications in fair machine learning and physics-informed learning are examined, demonstrating the performance of our proposed algorithms compared to alternative unconstrained and constrained optimization techniques.

3 - A Quasi-Newton Method for Nonsmooth, Nonconvex, Stochastic Optimization

Luke Marrinan, Pennsylvania State University, State College, PA, United States, Uday Shanbhag, Farzad Yousefian

We consider the minimization of a Lipschitz continuous and expectation-valued function over a closed and convex set. Our focus lies on obtaining both asymptotics as well as rate and complexity guarantees for computing an approximate stationary point (in a Clarke sense) via zeroth-order schemes. We adopt a randomized-smoothing-based approach reliant on minimizing a smooth approximation of our objective function. In such a setting, we develop a zeroth-order stochastic quasi-Newton scheme reliant on a combination of randomized and Moreau smoothing is proposed and analyzed, for which iteration and sample complexities are derived.

4 - A First-Order Augmented Lagrangian Method for Constrained Minimax Optimization

Sanyou Mei, University of Minnesota, Minneapolis, MN, United States, Zhaosong Lu

Minimax optimization with coupled constraints has received tremendous amount of attention in the recent years, finding widespread applications in many areas such as adversarial training, reinforcement learning and distributed computing. In this talk, we will introduce a first-order augmented Lagrangian method for solving a class of constrained minimax optimization with coupled constraints. Under some suitable assumptions, we will establish an operation complexity, measured by its fundamental operations, for the first-order augmented Lagrangian method for finding an approximate KKT solution of the constrained minimax problem.

SC35

Summit - 427

Learning in Transportation Network Modeling (Part 2)

Invited Session

TSL: Intelligent Transportation Systems

Chair: Minghui Wu, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Zhichen Liu, University of Michigan, Ann Arbor, United States

1 - Online Routing for Connected Vehicles Against Stealthy Cyberattacks

Minghui Wu, University of Michigan, Ann Arbor, MI, United States, Yafeng Yin, Jerome Lynch

This talk introduces a novel online routing algorithm, wherein a connected and automated vehicle recursively leverages real-time traffic predictions to make informed decisions. The algorithm is designed to be robust to cyberattacks on prediction services, which not only withstands adverse conditions but also maintains high efficiency in normal operational conditions. The proposed algorithm achieves this by decomposing decision-making across various nodes, with each node executing an individual sub-algorithm that considers both immediate and long-term effects on the routing process. We theoretically establish a sub-linear regret bound for the algorithm, offering a worst-case performance guarantee. When implemented in a practical routing scenario using real-world data, the algorithm demonstrates its effectiveness beyond merely adversarial contexts. Unlike traditional defense approaches that are designed against particular attack models, our approach requires no knowledge of underlying prediction mechanisms and attack strategies, paving the way for a more reliable and safe transportation system.

2 - Providing Real-TIME En-Route Suggestions to Cavs for Congestion Mitigation: a Two-Way Deep Reinforcement Learning Approach

Xiaoyu Ma, Rensselaer Polytechnic Institute, Troy, NY, United States, Xiaozheng He

This research investigates the effectiveness of information provision for congestion reduction in Connected Autonomous Vehicle (CAV) systems. The inherent advantages of CAVs, such as vehicle-to-everything communication, advanced vehicle autonomy, and reduced human involvement, make them conducive to achieving Correlated Equilibrium (CE). Leveraging these advantages, this research proposes a reinforcement learning framework involving CAVs and an information provider, where CAVs conduct real-time learning to minimize their individual travel time, while the information provider offers real-time route suggestions aiming to minimize the system's total travel time. The en-route routing problem of the CAVs is formulated as a Markov game and the information provision problem is formulated as a single-agent Markov decision process. Then, this research develops a customized two-way deep reinforcement learning approach to solve the interrelated problems, accounting for their unique characteristics. Moreover, CE has been formulated within the proposed framework. Theoretical analysis rigorously proves the realization of CE and that the proposed framework can effectively mitigate congestion without compromising individual user optimality. Numerical results demonstrate the effectiveness of this approach. Our research contributes to the advancement of congestion reduction strategies in CAV systems with the mitigation of the conflict between system-level and individual-level goals using CE as a theoretical foundation. The results highlight the potential of information provision in fostering coordination and correlation among CAVs, thereby enhancing traffic efficiency and achieving system-level goals in smart transportation.

3 - Online Relocating and Matching of Ride-Hailing Services: a Model-Based Modular Approach

Chang Gao, Department of Industrial Engineering, Tsinghua University, Beijing, China, People's Republic of, Xi Lin, Fang He, Xindi Tang

This study proposes an innovative model-based modular approach (MMA) to dynamically optimize order matching and vehicle relocation in a ride-hailing platform. MMA utilizes a two-layer and modular modeling structure. The upper layer determines the spatial transfer patterns of vehicle flow within the system to maximize the total revenue of the current and future stages. With the guidance provided by the upper layer, the lower layer performs rapid vehicle-to-order matching and vehicle relocation. MMA is interpretable, and equipped with the customized and polynomial-time algorithm, which, as an online order-matching and vehicle-relocation algorithm, can scale past thousands of vehicles. We theoretically prove that the proposed algorithm can achieve the global optimum in stylized networks, while the numerical experiments based on both the toy network and realistic dataset demonstrate that MMA is capable of achieving superior systematic performance compared to batch matching and reinforcement-learning based methods. Moreover, its modular and lightweight modeling structure further enables it to achieve a high level of robustness against demand variation while maintaining a relatively low computational cost.

4 - Discovering Traffic Dynamics from Trajectory Data Through Deep Learning

Ohay Angah, University of Washington, Seattle, WA, United States, Xuegang Ban

This study aims to discover traffic dynamics from vehicle trajectories through automatic learning techniques considering expression parsimony and interoperability. We propose an expression exploration framework based on deep symbolic regression (DSR) while encouraging parsimonious mathematical expressions to include potential variable interactions performed from data. In the expression exploration process, two penalty terms are added to the reward function to train recurrent neural networks (RNNs): (i) a complexity penalty to regulate the complexity of explored expressions and (ii) a variable interaction penalty to encourage the exploration to explore expressions consisting of potential feature combinations. Instead of limiting the expression complexities in exploration aggressively, the complexity penalty strategy is embedded in the reward function to encourage the RNNs to explore expressions with reasonable complexities. The strategy is found to be able to better regulate the complexities of expression candidates than the existing DSR framework. In the experiments, we further investigate the robustness of the proposed framework. Results demonstrate that our framework is efficient and robust for discovering mathematical representations of car-following dynamics. The proposed reward function performs well in regulating the complexity range of expression search.

SC36

Summit - 428

New Applications of Routing Problems

Invited Session

TSL: Freight Transportation

Chair: Rui Zhang, University of Colorado Boulder, Boulder, CO, 80309, United States

Co-Chair: Mengting Chao, University of Maryland-College Park, College Park, MD, United States

1 - On-Demand Routing for B2b Retailers

Mengting Chao, University of Maryland-College Park, College Park, MD, United States

COVID-19 significantly changed the last-mile delivery network for B2B retailers. Due to work-from-home, delivery locations were more spread out in the suburbs with 1-2 boxes per stop compared to a larger number of boxes in more centralized office locations. This exposed a general underlying problem in last-mile delivery when demand (total route time) exceeds delivery capacity (temporal). The challenge is to deliver to all customers on the promised delivery day with the retailer driver (RD) staying within the regular shift hours. Evolving industry practices include outsourcing some deliveries to on-demand drivers (ODDs). We aim to determine the route of the RD, the locations that the ODDs will deliver, and the drop-off locations where the RD will hand over packages to the ODDs.

2 - Disaster Response on a Network with Stochastic Demand and Uncertain Edge Accessibility

Jessa Rhea, University of Iowa - Applied Mathematical and Computational Sciences, Iowa City, IA, United States, Jeffrey Ohlmann

We present a mobile facility routing problem to provide safe and immediate disaster response for a network in which there is uncertain edge traversability and stochastic beneficiary demand. Our problem is formulated as a Markov decision process, and the objective is to maximize the amount of beneficiary demand served by the capacitated fleet over the problem horizon. We note that a mobile facility can contribute to this objective even if it does not serve demand at an epoch (e.g., when its capacity is entirely depleted) by determining accessibility of edges and observing beneficiary demand at locations along its route. Our solution approach is a reinforcement learning algorithm benchmarked by the vehicle routing problem with stochastic demand.

3 - The Truck-and-Robot Routing Problem with Pickups and Deliveries

Manuel Ostermeier, University of Augsburg, Augsburg, Germany, Tobias Huf

The increasing number of home deliveries paired with various delivery modes and channels require retailers to establish efficient last-mile solutions. The innovative truck-and-robot concept poses a promising approach in this regard. The concept relies on Sidewalk Autonomous Delivery Robots (SADRs) carried and released by trucks to serve customers in predefined time windows. We extend existing concepts by integrating store locations and corresponding customer requests for direct deliveries from stores. This setup combines the first and the last mile, giving rise to a novel concept where different delivery modes are established that provide robots as a service to pick up and deliver goods to customers.

The described problem is formalized as the Truck-and-Robot Pickup-and-Delivery Problem (TnR-PDP) and solved using a tailored approach, the Adaptive Genetic Algorithm (AGA). The AGA is based on a recombination-based search framework but tailored to the problem-specifics (e.g., no or multiple visits per location) using specialized recombination operators and an adaptive search strategy for location and operator selection.

The numerical experiments show that our approach outperforms existing benchmark approaches concerning runtime by up to 79 % while also improving the solution quality. The AGA solves relevant instances in a reasonable amount of time and generates robust solutions. Further analyses highlight the value of store integration into the truck-and-robot concept and show the benefits of the TnR-PDP compared with commonly applied approaches in the industry.

4 - Estimating Optimal Solution Values in TSP, VRP, and SDVRP

Shuhan Kou, University of Maryland, College Park, College Park, MD, United States, Bruce Golden, Luca Bertazzi

Motivated by Basel and Willemain's (2001) work on estimating optimal tour lengths in the Traveling Salesman Problem (TSP), we identify an intrinsic relationship between the distribution of the feasible solution space and the optimal solution value in combinatorial optimization problems. We begin with a linear regression model that uses standard deviation as a predictor to predict the optimal TSP tour length. Building on this, we improve our model by incorporating the mean and minimum predictors, significantly enhancing predictive accuracy. We also extend this method to over 10,000 vehicle routing problem (VRP) instances. To tackle the more complex split delivery vehicle routing problem (SDVRP), we integrate mean, standard deviation, and two other topological features into our model. Evaluating our approach across 95 diverse benchmark instances, our regression model provides an average margin of 3%. Moreover, our recent experiments reveal that employing these predictors in random forest and neural network models further enhances the predictive accuracy beyond that of linear regression for the TSP and the VRP.

5 - An empirical study of a compact formulation of the capacitated fixed-charge network flow problem

Eli Olinick, Southern Methodist University, Dallas, TX, United States

The triples formulation is a compact formulation of multicommodity network flow that provides an alternative to the traditional and widely used node-arc and arc-path representations. When applied in mixed integer programming formulations of the capacitated fixed charge network flow problem (CFCNFP), the triples formulation significantly reduces the number of constraints and continuous variables compared to the standard node-arc formulations. In an empirical study, we find that CFCNFP instances that can be solved to provable optimality are solved an average of 2.75 times faster with the triples formulation than with the node-arc formulation. For instances that cannot be solved to provable optimality in our computing environment, we find that given a 30-minute time limit the triples formulation produces solutions with lower costs and optimality gaps.

SC37

Summit - 429

Transportation and Optimization for Strategic Decisions and Policymaking

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Arthur Delarue, Georgia Institute of Technology, Atlanta, GA, United States

1 - School District Design Under Uncertainty

Aysu Ozel, Northwestern University, Evanston, IL, United States, Karen Smilowitz

Many decisions in school district design aim to provide continuity in educational experiences, particularly long-term district design decisions that include opening/closing schools, drawing attendance boundaries, and placing programs within schools. Given the need for continuity, it is important to consider the robustness of district design options with respect to uncertain changes. We explore school district design under uncertainty in the larger context of decision making under uncertainty, considering the unique features of school settings. We examine multiple sources of uncertainty in school enrollment, including exogenous demographic changes and endogenous changes stemming from community response to district design updates. Further, we consider modeling of decisions given different levers to address uncertain enrollment and current practices in school districts.

2 - On-trip matching in shared rides

Julia Yan, University of British Columbia, Vancouver, BC, Canada, Yifan Shen, Chiwei Yan

We will discuss issues related to on-trip matching in shared rides.

3 - A Scalable Algorithm for Large-Scale Network Design with Service Time Guarantees

Myungeun Eom, Georgia Institute of Technology, Atlanta, GA, United States, Alejandro Toriello, Alan Erera

We propose a scalable algorithm for a large-scale network design problem with service time guarantees, motivated by small package delivery systems. Our model captures an important constraint that commodities must be delivered within specified service time periods. We aim to determine paths for commodities to minimize transportation and hold costs while ensuring the service time guarantees. We solve this problem on the entire United States network, which includes approximately 250 hub buildings and 40,000 commodities. The main challenges of this problem arise from the large network size and the high volume of commodities. To address the scalability issue from the network size, we partition the network into smaller sub-networks considering geographic features. We solve each sub-network independently for commodities whose origins and destinations are within the same sub-network. For commodities with origins and destinations in different sub-networks, we identify networks containing potential transfer terminals and solve the problem within these chosen networks. To handle a large number of commodities, we develop a batching method that prioritizes commodities with less flexibility. We iteratively solve the problem for each batch, building on the previous solutions for consolidation. We demonstrate the effectiveness of our approach through a computational study conducted on networks of various sizes.

4 - Two-stage stochastic optimization formulation for the police staffing problem

Lynn Xu, Georgia Institute of Technology, Atlanta, GA, United States

We present a two-stage stochastic optimization formulation of the police staffing problem, where staffing decisions are made in the first stage and dispatch decisions in the second stage. Noticing the special structure of our problem, we provide a customized integer L-shaped method for our second-stage problem and show its computational advantage compared to the existing methods using 911 call data from the Atlanta Police Department. We also develop a new formulation for the second-stage problem, which adopts ideas from interval scheduling theory.

5 - Novel Regret Bounds based on Information Relaxations

Lavanya Marla, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Tharunkumar Amirthalingam

We present a novel approach to evaluating policies for large-scale resource allocation problems on networks. Strategic and tactical policies are often evaluated based on regret bounds, which may be too loose. Our approach is based on the idea of penalized information-relaxation bounds, proposed earlier by Brown, Smith and Sun (2010). We identify and solve methodological gaps in the existing approach and demonstrate how these can be successfully operationalized for large-scale decision-making problems on networks. We demonstrate our methodology on applications involving airline recovery, emergency medical service policies and training autonomous vehicles.

SC38

Summit - 430

AAS Student Presentation Competition I

Award Session

Aviation Applications

Chair: Wayne Ng, Singapore University of Technology and Design, Singapore, Singapore

1 - Airline competitive Hub-and-Spoke network design by reweighted norm-1

Fernando Real Rojas, King Juan Carlos University, Mostoles, Spain, Luis Cadarso, Antonio Marques

The widespread adoption of hub-and-spoke networks in the airline sector has sparked interest in hub location problems. Our research introduces a profit-maximizing, flow-based mathematical model tailored for designing airlines' capacitated networks. Unlike traditional methods assuming a predefined hub-and-spoke structure, we allow hubs to emerge based on cost-effectiveness, passenger preferences, and airline competition.

Given the competitive nature of the airline industry, we focus on optimizing network design while considering passenger behavior. Discrete choice models capture factors influencing passenger decisions, such as route price and travel time, integrating supply and demand interactions. These models, like the logit model, are nonlinear and nonconvex, posing solver development challenges. We address this with a reverse engineering approach, adding convex negative entropy passenger flow regularizers, leading to logit-based solutions.

Our formulation jointly considers key decision variables like airport base locations, route offerings, and associated capacities. This complicates the optimization but integrates network infrastructure design with passenger choices and market shares. To solve this, we propose a novel approach using convex relaxation and reweighted norm-1 minimization, combining decision and dimensioning network infrastructure variables, and considering operating costs and passenger decisions.

A strategic element of our approach is tailored flow conservation constraints that only allow transfers at hubs, crucial for yielding hub-and-spoke topologies and modeling passenger choices. Our method optimizes hub-and-spoke network design considering profitability, demand forecasts, passenger choices, and competitor presence. We demonstrate effectiveness with a case study of a new domestic airline across 25 U.S. airports, validating the methodology and offering insights into the hub-and-spoke design.

2 - Award Session Speaker

3 - Award Session Speaker

4 - Learning for Collaborative Aircraft Trajectory Optimization in Airspace

5 - Accelerated Column Generation approach for Tail Assignment Problem

SC39

Summit - 431

Railway Timetabling and Demand Forecasting

Invited Session

Railway Applications

Chair: Pengli Mo, N/A, United States

Co-Chair: Marcella Samà, Roma Tre University, Roma, Italy

1 - Delay risk evaluation in real-time train rescheduling under uncertain dwell times

Marcella Samà, Roma Tre University, Roma, Italy, Marco Pranzo, Carlo Meloni

Given a train timetable made infeasible by unexpected disturbances, the real-time Railway Traffic Management Problem (rtRTMP) returns a new feasible plan of operations in which corrective decisions have been taken to minimize delay propagation. The rtRTMP is a very well-studied problem in the literature. Usually, given the necessity to quickly return a good quality solution, the problem is solved considering deterministic passengers dwell time, disregarding how uncertainty may affect the new constructed plans. Such simplification of the problem may lead to unexpected worsening of the performance promised by an rtRTMP solver, which in turn may negatively impact the confidence of dispatchers on such tools. A clearer picture of the actual rtRTMP solution quality, especially in case of worst-case scenarios realizations, would be preferable, to better allow dispatchers to take more informed decisions. In this talk we present an approach to assess the risk of worsening delays associated to an rtRTMP solution when dwell times are uncertain. We explore the evaluation of the Conditional-Value-at-Risk of the maximum train delay as risk index. We model an rtRTMP solution as a temporal activity network and apply an innovative numerical method to obtain a real-time risk evaluation for rtRTMP solutions. We test the proposed approach on a real case study analyzing the effects of uncertainties of different severity. Different computational schemes are presented, testing and analyzing their suitability for practical contexts requiring risk-averse or risk-aware scheduling processes. The methodology appears to be promising to deal with the different risk attitudes of the decision-makers.

2 - A Fragility-based Approach to Timetable Design

Marta Leonina Tessitore, Roma Tre University, Rome, Italy, Marcella Samà, Giorgio Sartor, CARLO MANNINO, Dario Pacciarelli

In a typical tactical timetabling process, route planners are usually tasked to produce a timetable for the next months or year. They start from an existing timetable and follow a time-consuming trial-and-error process to obtain a new timetable. Given the increase in traffic demand, public service companies require the implementation of effective decision support systems that can guide practitioners throughout their decision-making process, regardless of their experience level. In this work, we present a fragility-based approach to timetable design. Specifically, we show how the concept of fragility can be easily exploited by route planners to design new and more robust timetables, i.e., to suggest where to add new time supplements or redistribute the available ones. We propose a timetabling model that aims at building a more robust timetable, modifying the existing timetable, and reducing the fragility of its most critical sections, that is a single large model capable of simultaneously doing timetabling and dispatching. We solve this model using a delayed row generation algorithm similar to the one first described in Lamorgese and Mannino (2015). Each iteration of the algorithm produces a new timetable that may contain different fragile sections. The process can be iterated until the desired level of robustness is achieved. Considering real-life scenarios from a Norwegian railway line, we show that we can significantly improve the fragility of a timetable, even when considering conservative strategies for possible improvements.

3 - Collaborative Optimization of Rolling Stock Allocation and Timetable Coordination in a Multi-Modal Rail Network: MILP Formulation and Decomposition-Based Algorithm

Jiateng Yin, Beijing Jiaotong University, Beijing, China, People's Republic of

The transportation system is shifting towards a shared mobility ecosystem, emphasizing resource aggregation and coordination among different transport modes to enhance service quality and encourage passengers to shift from private cars to public transit. This study focuses on the coordination between metro and rail networks in a city, to alleviate congestion at transfer hubs and to decrease the travel time of passengers. Given the "tide-like" transfer passengers between rail and metro hubs, we investigate the potential of reserving a number of rolling stocks, strategically allocated to the metro trains with the highest demand to avoid over-congestion. We present a mixed integer linear programming (MILP) model to formulate this problem, with decision variables including rolling stock allocation and coordinated schedules for both metro and rail networks, by considering the time-varying demand of passengers. The objectives aim to minimize the passenger travel/transfer time and operational costs for managers. Then, we analyze the mathematical properties and propose an exact decomposition-based solution algorithm. Our algorithm reformulates the original problem into a relaxed master problem (corresponding to rolling stock allocation and timetabling) and a series of independent subproblems (corresponding to passenger flows). We provide optimality conditions and prove that the subproblems can be solved by an analytical procedure. According to these properties, we further propose a set of feasibility cuts and we prove that the new cuts provide tighter bounds in comparison with traditional Benders cuts. We test our integrated approach and solution algorithms on real-world instances from the Beijing railway network.

4 - Passenger-oriented rolling stock scheduling in the metro system with multiple depots: Network flow based approaches

Entai Wang, HEC Montreal, Montreal, QC, Canada, Lixing Yang, Jiateng Yin

This study investigates a rolling stock scheduling problem on a metro line with multiple depots. Two novel optimization models, i.e., an arc-based and a path-based network-flow models, are formulated with the aim of improving the service level and reducing the operation cost simultaneously, in which the flexible train composition mode is also taken into consideration to well match the transport capacity and time-varying passenger demand. To solve the proposed models, a branch-and-price (B&P) approach is designed to find the near optimal operation schemes, in which the column generation is used to solve the relaxed problem at each node of the searching tree, where a dynamic programming approach is embedded to solve the pricing sub-problem associated with each depot to generate promising paths (columns) for each rolling stock unit, and then the branch-and-bound (B&B) procedure is incorporated to find integral solutions. To test the performance of the proposed approaches, a series of numerical experiments are conducted both on small-scale and real-life cases of the Beijing metro Batong line with historically recorded passenger data. The computation results have verified the improved operational efficiency and a better service level of the solutions found by our proposed approaches.

SC40

Summit - 432

Human-in-the-loop in Forecast and Inventory Systems

Invited Session

Behavioral Operations Management

Chair: Enno Siemsen, University of Wisconsin-Madison, Madison, WI, United States

1 - Inventory and Forecast Adjustments

Eirini Spiliotopoulou, Tilburg University, Tilburg, Netherlands, Enno Siemsen

We examine whether planners should be allowed to adjust forecasts or the resulting inventory orders (or both). We use experiments to establish that order adjustments work better due to less noise in the decisions. Several experiments examine the behavioral cause for this phenomenon, and compare students and professional forecasters as participants in the experiment.

2 - Reluctant to Decrease: Human-Algorithm Collaboration in Tactical Inventory Decisions

Joan Stip, Eindhoven University of Technology, Eindhoven, Netherlands, Jan Fransoo, Geert-Jan van Houtum

We study augmented decision-making when setting base stock levels in high-stakes spare parts inventory management. Such spare parts inventory management is characterized by many 0-1 decisions: the base stock level is either 0 or 1. We partner with an OEM where state-of-the-art inventory control algorithms have been deployed and observe that human decision-makers deviate from the optimal solution by adjusting 24% of the base stock levels. We deploy a structural estimation model estimating the humans' psychological cost for overage and underage and find that in particular, the psychological cost for underage are substantially higher than parameterized in the algorithms. Interestingly, we can show that such psychological costs for underage are particularly high in case the algorithm proposes to reduce the inventory from 1 to 0, similar to the well-known endowment effect. Leveraging our structural estimation, we run a counterfactual analysis and show that a Pareto-optimal alternative solution can be obtained that is more aligned with the humans' perceptions of underage costs.

3 - Navigating Forecast Dynamics: the Decision Dilemma of Multi-Ordering Newsvendors

Michael Becker-Peth, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Anne Dohmen, Stephanie Eckherd, Lance Saunders

In supply chain management, orders are typically placed well in advance based on forecasts. However, these forecasts may evolve over time, leading to potential inaccuracies in the original orders. This study investigates the decision-making process of human operators facing the option to update order quantities in a multi-ordering newsvendor system. Through controlled laboratory experiments, we examine how variations in mean demand, demand uncertainty, and change fees influence the propensity to adjust initial orders. Our findings reveal that changes in mean demand consistently trigger order updates, regardless of the potential benefits, while adjustments in demand uncertainty result in fewer updates. This dynamic introduces the risk of both false positives—updating unnecessarily—and false negatives—failing to update when beneficial—impacting overall system performance. Thus, understanding the underlying decision mechanisms is crucial for determining the impact of providing the option to update orders. Our research sheds light on the complexities of navigating forecast dynamics in multi-ordering newsvendor systems, highlighting the balance between the potential gains from order adjustments and the associated risks.

4 - Profit Implications of Judgmental Adjustments to Forecast Inputs: Evidence from a Large-Scale Field Experiment

Dayton Steele, University of Minnesota, Carlson School of Management, Minneapolis, MN, United States, Saravanan Kesavan, Tarun Kushwaha

In this paper, we report the results from a large-scale field experiment at a spare parts retail chain that considers whether allowing merchants to override forecast inputs to an inventory algorithm improves profits. While the judgmental forecasting literature has studied extensively whether judgmental adjustments improve forecast performance, causal empirical evidence is missing in regards to whether judgmental adjustments improve bottom-line profits. Our results show that judgmental adjustments to the forecast input increase profitability by 3.98% on average compared to relying on automation without human intervention. We find that the well-established motivation-opportunity-ability framework provides clear insight into when judgmental adjustments improve profits, by examining heterogeneity in our data regarding SKU margin, lifecycle, and size of supplier. Our dataset also allows for examining both forecast accuracy and profits. We empirically support the wisdom from the judgmental forecasting literature that forecast improvements need not translate to profit improvements, calling attention to the need to consider operational performance beyond forecast accuracy as an end in itself.

SC41

Summit - 433

Causal Inference & ML

Invited Session

Applied Probability Society

Chair: Kelly Zhang, Columbia Business School (DRO), New York, NY, United States

1 - On Experiment Design Under Network Interference

Mohsen Bayati, Stanford University, Standford, CA, United States, Yuwei Luo, William Overman, Sadegh Shirani, Ruoxuan Xiong

Randomized experiments are a powerful methodology for data-driven evaluation of decisions or interventions. Yet, their validity may be undermined by network interference. This occurs when the treatment of one unit impacts not only its outcome but also that of connected units, biasing traditional treatment effect estimations. This talk discusses a new framework to accommodate complex and unknown network interference and produce practical estimators.

2 - Minimax Optimal Estimates of Individual Causal Effects in Panel Data under General Intervention Patterns

Christina Yu, Cornell University, Ithaca, NY, United States, Yudong Chen, Xumei Xu

Consider estimating individual causal effects of a treatment on an individual i at time t from panel data, where we observe the outcomes of multiple units across a period of time. Each unit may be exposed to the treatment at different time points across the horizon. Assuming the two-way linear fixed effects model, for any general intervention pattern, we present tight conditions for identification of individual causal effects along with a simple unbiased estimator. The estimator is both minimax optimal and is a uniform minimum variance unbiased estimator when the observation noise is independent from the treatment variables. In particular, the minimax achievable squared error for recovery of the causal effect associated to the (i,t) unit-time pair is proportional to the effective resistance of (i,t) in the bipartite graph associated to the intervention pattern, and the estimator is derived from the unit (i,t) electrical flow on that graph. We show that our estimator can be viewed as a generalization of the difference in differences estimator, and our results provide insights for understanding two-way fixed effects regression estimators.

3 - Adaptive Experimentation at Scale: A Mathematical Programming Approach

Hongseok Namkoong, Columbia University, New York, NY, United States, Ethan Che

Adaptive experimentation can improve statistical power significantly, but typical algorithms overlook important issues that arise in practice: multiple objectives, non-stationarity, batched/delayed feedback, constraints, and personalization. Moving away from developing bespoke algorithms for each setting, we present a mathematical programming view of adaptive experimentation that can flexibly incorporate a wide range of objectives, constraints, and statistical procedures. By formulating a dynamic program in the batched limit, our modeling framework enables the use of scalable optimization methods (e.g., SGD and auto-differentiation) to solve for treatment allocations. To spur algorithmic progress, we build a suite of benchmark problems based on hundreds of real A/B tests at ASOS that model key practical issues such as non-stationarity, personalization, multi-objectives, and constraints. Our empirical results show standard Thompson sampling-based policies fail to reliably improve upon static designs, and demonstrate the effectiveness of a simple planning approach.

4 - Orthogonal Estimation of the Difference-of-Q-Functions

Angela Zhou, USC Marshall School of Business Data Sciences and Operations, Los Angeles, CA, United States

Offline reinforcement learning is important in many settings with available observational data but the inability to deploy new policies online due to safety, cost, and other concerns. Many recent advances in causal inference and machine learning target estimation of heterogeneous treatment effects, and more broadly causal contrast functions. Estimating causal contrasts is sufficient for making optimal decisions and can adapt to potentially smoother structure. We develop a dynamic generalization of the R-learner for estimating the difference of Q-functions, Q(s,1)-Q(s,0) (and for multiple actions, fixing a choice of a_0, Q(s,a)-Q(s,a_0). The method wraps around standard estimation procedures in offline reinforcement learning via a sequence of sequential loss minimization problems, which makes it appealingly practical. But, the residualized learning step allows for targeting the more structured smoothness of the Q-function contrast itself. We illustrate with relevant examples, including "decision-theoretic sparsity" of the difference-of-Q function that arises due to joint structure of reward and transitions.

SC42

Summit - 434

Topics in Data-driven Stochastic Optimal Control

Invited Session

Applied Probability Society

Chair: Andrew Lim, National University of Singapore, Singapore, Singapore

Co-Chair: Wang Chi Cheung, National University of Singapore, Singapore, Singapore

1 - Q-Learning in Continuous TIME

Xunyu Zhou, Columbia University, New York, NY, United States, Yanwei Jia

We study the continuous-time counterpart of Q-learning for reinforcement learning (RL) under the entropy-regularized, exploratory diffusion process formulation introduced by Wang et al. (2020). As the conventional (big) Q-function collapses in continuous time, we consider its first-order approximation and coin the term “(little) q-function”. This function is related to the instantaneous advantage rate function as well as the Hamiltonian. We develop a “q-learning” theory around the q-function that is independent of time discretization. Given a stochastic policy, we jointly characterize the associated q-function and value function by martingale conditions of certain stochastic processes, in both on-policy and off-policy settings. We then apply the theory to devise different actor–critic algorithms for solving underlying RL problems, depending on whether or not the density function of the Gibbs measure generated from the q-function can be computed explicitly. One of our algorithms interprets the well-known Q-learning algorithm SARSA, and another recovers a policy gradient (PG) based continuous-time algorithm proposed in Jia and Zhou (2022b). Finally, we conduct simulation experiments to compare the performance of our algorithms with those of PG-based algorithms in Jia and Zhou (2022b) and time-discretized conventional Q-learning algorithms.

2 - Optimal Trade Execution with Learning

Andrew Lim, National University of Singapore, Singapore, Singapore, Galvin Ng

We consider the problem of optimal trade execution with a linear price impact function when both the drift (intercept) and permanent price impact (slope) parameters are uncertain. Since the trading rate affects the returns data that is used to learn the uncertain parameters, this problem is intrinsically a tradeoff between exploration and exploitation, and it is of particular interest to understand how the trader goes about "exploring" under the optimal policy. To sidestep the well-known challenges of solving the dynamic programming equation, which is infinite dimensional in general and has no explicit solution even when the prior on both parameters is Gaussian, we consider the asymptotic regime where the learning rate vanishes and use a Taylor Series expansion extract the exploration component of the optimal policy. This exploration term can be easily characterized and computed, and provides surprising but ultimately intuitive insights into the nature of exploration, its dependence on the time horizon, the inventory level and the value of learning the drift and market impact parameters.

3 - Leveraging (Biased) Information: Multi-armed Bandits with Offline Data

Lixing Lyu, National University of Singapore, Singapore, Singapore, Wang Chi Cheung

We leverage offline data to facilitate online learning in stochastic multi-armed bandits. The probability distributions that govern the offline data and the online rewards can be different. Without any non-trivial upper bound on their difference, we show that no non-anticipatory policy can out-perform the UCB policy by (Auer et al. 2002), even in the presence of offline data. In complement, we propose an online policy MIN-UCB, which outperforms UCB when a non-trivial upper bound is given. MIN-UCB adaptively chooses to utilize the offline data when they are deemed informative, and to ignore them otherwise. MIN-UCB is shown to be tight in terms of both instance independent and dependent regret bounds. Finally, we corroborate the theoretical results with numerical experiments.

4 - Continuous Time Reinforcement Learning using Rough Paths

Prakash Chakraborty, The Pennsylvania State University, University park, PA, United States, Harsha Honnappa, Samy Tindel

This study delves into an optimal control formulation for continuous time reinforcement learning in rough environments. To model uncertainty and facilitate effective exploration of the environment, our approach incorporates relaxed controls into the controller's actions. We carefully study the properties of the corresponding relaxed and rough Hamilton-Jacobi-Bellman (HJB) equation and discuss how our rough control framework can enhance reinforcement learning strategies.

SC43

Summit - 435

Learning in Changing Environments

Invited Session

Applied Probability Society

Chair: Gal Mendelson, Technion, Haifa, Israel

1 - Roping in Uncertainty: Robustness and Regularization in Markov Games

Qiaomin Xie, University of Wisconsin-Madison, Madison, WI, United States, Jeremy McMahan, Giovanni Artiglio

We study robust Markov games (RMG) with s-rectangular uncertainty. We show a general equivalence between computing a robust Nash equilibrium (RNE) of a s-rectangular RMG and computing a NE of an appropriately constructed regularized MG. The equivalence result yields both a planning algorithm for solving s-rectangular RMGs and provable robustness guarantees for policies computed using regularized methods. However, we show that even for just reward-uncertain two-player zero-sum matrix games, computing an RNE is PPAD-hard. Consequently, we derive a special uncertainty structure called efficient player-decomposability, and show that RNE for two-player zero-sum RMG in this class can be provably solved in polynomial time. This class includes commonly used uncertainty sets such as L_1 and L_\infty ball uncertainty sets.

2 - A Definition of Non-Stationary Bandits

Yueyang Liu, Stanford University, Stanford, CA, United States, Benjamin Van Roy, Kuang Xu

Despite the subject of non-stationary bandit learning having attracted much recent attention, we have yet to identify a formal definition of non-stationarity that can consistently distinguish non-stationary bandits from stationary ones. Prior work has characterized non-stationary bandits as bandits for which the reward distribution changes over time. We demonstrate that this definition can ambiguously classify the same bandit as both stationary and non-stationary; this ambiguity arises in the existing definition's dependence on the latent sequence of reward distributions. Moreover, the definition has given rise to two widely used notions of regret: the dynamic regret and the weak regret. These notions are not indicative of qualitative agent performance in some bandits. Additionally, this definition of non-stationary bandits has led to the design of agents that explore excessively. We introduce a formal definition of non-stationary bandits that resolves these issues. Our new definition provides a unified approach, applicable seamlessly to both Bayesian and frequentist formulations of bandits. Furthermore, our definition ensures consistent classification of two bandits offering agents indistinguishable experiences, categorizing them as either both stationary or both non-stationary. This advancement provides a more robust framework for agent design and analysis in non-stationary bandit learning.

3 - Average Reward Restless Bandits: Unichain and Aperiodicity are Sufficient for Asymptotic Optimality

Yige Hong, Carnegie Mellon University, Pittsburgh, PA, United States, Qiaomin Xie, Yudong Chen, Weina Wang

We consider the infinite-horizon, average-reward restless bandit problem in discrete time. We propose a new class of policies that are designed to drive a progressively larger subset of arms toward the optimal distribution. We show that our policies are asymptotically optimal with an $O(1/\sqrt{N})$ optimality gap for an N-armed problem, provided that the single-armed relaxed problem is unichain and aperiodic.

Our approach departs from most existing work that focuses on index or priority policies, which rely on the Uniform Global Attractor Property (UGAP) to guarantee convergence to the optimum, or a recently developed simulation-based policy, which requires a Synchronization Assumption (SA).

4 - An Information-Theoretic Analysis of Nonstationary Bandit Learning

Seungki Min, KAIST, Daejeon, Korea, Republic of, Daniel Russo

In nonstationary bandit learning problems, the decision-maker must continually gather information and adapt their action selection as the latent state of the environment evolves. In each time period, some latent optimal action maximizes expected reward under the environment state. We view the optimal action sequence as a stochastic process, and take an information-theoretic approach to analyze attainable performance. We bound per-period regret in terms of the entropy rate of the optimal action process. The bound applies to a wide array of problems studied in the literature and reflects the problem's information structure through its information-ratio.

SC44

Summit - 436

Active Learning for Optimization

Invited Session

Simulation Society

Chair: Giulia Pedrielli, Arizona State University, Tempe, AZ, United States

1 - Recent Advances in Grey-Box Bayesian Optimization

Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

Bayesian optimization (BO) is a framework for global optimization of expensive-to-evaluate objective functions. Classical BO methods assume the objective function is a "black box". However, information about the objective function is often available. Recently, "grey-box" BO methods leveraging such information have been shown to provide dramatic performance improvements in a broad range of problems. In this talk, I will discuss new advances in grey-box BO and the applications these methods have unlocked.

2 - Fast and Effective Gaussian Process Based Bayesian Optimization Without Maximum Likelihood Estimation

Antonio Candelieri, University of Milano-Bicocca, Milano, Italy

A novel Gaussian Process based Bayesian Optimization framework is presented, in which the modelling of the objective function and the selection of the new incumbent solution are simultaneously addressed by predicting the most suitable value of the Gaussian Process's kernel kyperparameter(s), avoding expensive computation of the Maximum Likelihood Estimation. Indeed, the Gaussian Process model is fitted at a constant time, independently on the number of function evaluations performed so far. At the same time, the prediction of the most suitable value of the kernel's hyperparameter(s) also avoids instability in the kernel matrix. As a result, the approach can be applied to globally optimize noise-free functions, without the need to add artificial noise (also known as nugget effect) to avoid ill-conditioning.

Results on a set of diversified test problems empirically show the benefits of the proposed approach, in terms of effectiveness and efficiency, against the well known *no-regret* Gaussian Process Lower Confidence Bound. Specifically, the proposed approach provides a lower cumulative regret until the best observed value can be improved. Otherwise, the regret increases because exploration is intrinsically triggered by the impossibility to further improve, close to the current best solution, without incurring into ill-conditioning. This leads to a completely new way to learn the Gaussian Process by simultaneously considering risk for ill-conditioning and exploration-exploitation balance provided by the next incumbent solution.

3 - Stochastic Linear Bandits with Partial Observability

Gautam Dasarathy, Arizona State University, Tempe, AZ, United States, Vineet Gattani, Lalit Jain

We consider the problem of stochastic linear bandits where the decision vectors available to the agent may only be partially observed. This models several modern sequential decision-making applications, such as recommendation systems, where it is impossible, impractical, or even imprudent to fully observe these decision vectors. As one example, the recent migration from third-party cookies means that advertisers only get limited observability into the user context. In this work, we propose Partially Observed Linear Bandit (POLB), an algorithm for tackling this problem. Our method sequentially builds an estimate of the (possibly low-dimensional) subspace the decision vectors lie in, and then leveraging this estimate in a sequential decision making framework. Our theory elucidates the effect of the amount of observability and the presence of low-dimensional structure underlying the decision vectors on the regret performance of POLB. We complement this with empirical results that highlight the advantages of the proposed framework.

4 - Respecting the limit: Bayesian optimization with a known bound on the optimal value

Matthias Poloczek, Amazon, San Francisco, CA, United States

Bayesian optimization (BO) has become a powerful method for the sample-efficient optimization of expensive black-box functions that arise in materials discovery, hardware design, AutoML, or portfolio optimization, for example. These functions do not have a closed-form and are evaluated by running a complex simulation or an experiment. In many real-world optimization problems, we have prior information about what values are achievable under the objective function. We study the scenario that we have either exact knowledge of the value of the global minimum or a, possibly inexact, lower bound on its value. We propose Bound-Aware Bayesian Optimization (BABO) that uses a tailored surrogate model called SlogGP and a new acquisition criterion based on Expected Improvement to leverage the bound information. Empirical results on a variety of benchmarks demonstrate the benefit of taking prior information about the optimal value into account. Interestingly, we notice that even in the absence of such prior information, the new SlogGP surrogate model outperforms the standard GP model in many cases.

Joint work of Hanyang Wang, Juergen Branke (both Warwick Business School, UK), and Matthias Poloczek (Amazon, USA).

5 - Simulation-based digital twins: Identifying optimal alternatives using real-time data

Moones Keshvarinia, Iowa State University, Ames, IA, United States, Cameron MacKenzie

Digital twins have the potential to support decision making in real time in a manufacturing facility. The digital twin should run in parallel with the manufacturing facility so that the digital twin represents the current state of the manufacturing system. Simulation within the digital twin can be used to explore different scenarios and forecast potential outcomes in the manufacturing system. When problems arise in the manufacturing system, the manufacturer can use the digital twin simulation to identify alternatives to resolve these problems while incorporating real-time data from the manufacturing facility. This study investigates the use of a digital twin to resolve a severe supply shortage in a manufacturing facility. We use a discrete-event simulation to determine if a digital twin relying on real-time data from the manufacturing facility generates a different set of optimal alternatives that mitigate the effects of the supply shortage compared to the optimal alternatives generated from a static simulation without access to real-time data. The simulation models different operational alternatives such as adding more forklifts, employing condition-based maintenance, hiring more workers, and moving workers from one station to another station. Machine learning algorithms identify the combination of alternatives that maximizes the manufacturer's expected profit using data generated by the simulation. The findings from this research quantify the potential benefit of digital twins in manufacturing for the purpose of managing disruptions.

SC45

Summit - 437

Editorial Insights: A Panel Discussion with Healthcare Journal Editors

Panel Session

Health Applications Society

Co-Chair: Nan Liu, Boston College, Chestnut Hill, MA, United States

1 - Editorial Insights: A Panel Discussion with Healthcare Journal Editors

Nan Liu, Boston College, Chestnut Hill, MA, United States

2 - Panelist

Timothy Chan, University of Toronto, Toronto, ON, Canada

3 - Panelist

Diwakar Gupta, University of Texas, Austin, TX, United States

4 - Panelist

Lauren Lu, Dartmouth College, Hanover, NH, United States

5 - Panelist

Nicos Savva, London Business School, London, United Kingdom

6 - Panelist

Greg Zaric, Ivey Business School, London, ON, Canada

SC46

Summit - 438

Health Applications in Preventive Medicine

Invited Session

Health Applications Society

Chair: Osman Ozaltin, North Carolina State University, Raleigh, NC, United States

Co-Chair: Maria Mayorga, North Carolina State University, Raleigh

Co-Chair: Zhaowei She, Singapore Management University, Singapore, Singapore

1 - Urban-Rural Differences in Emergency Medical Services TIME Intervals for Traumatic Injury in the United States

Maria Mayorga, North Carolina State University, Raleigh, NC, United States

Purpose: This study aims to evaluate urban-rural differences in emergency medical services (EMS) time intervals for traumatic injury patients in the United States. We also explored the intersectionality of rurality and minority race/ethnicity on EMS time intervals.

Methods: We conducted a retrospective cohort study of EMS incidents for traumatic injury in the 2021 National EMS Information System (NEMSIS) Public Release Research Dataset. Prehospital time intervals, including EMS response, scene, and transport times, were analyzed by urban, suburban, and rural areas for the entire study population and within race/ethnicity subgroups. Multivariable quantile regression was used to estimate adjusted differences in median and 90th percentile times.

Findings: We analyzed 3,952,030 eligible EMS incidents for traumatic injury. Median EMS response, scene, and transport times were 7.1, 14.7, and 12.0 minutes, respectively. In adjusted quantile regression analyses, urban areas had the shortest median and 90th percentile response and scene times while suburban areas had the shortest transport times. Statistically significant differences in EMS time intervals were observed between race/ethnicity groups within each urbanicity level.

Conclusions: In a national database of EMS encounters, EMS response and transport times for traumatic injuries occurring in rural areas were substantially longer compared to urban areas, adjusting for patient demographics and other incident-level factors. Further research on the impact of prehospital emergency care on injury-related disparities in rural and other underserved populations is needed.

2 - Optimizing HPV Intervention Strategies: A Reinforcement Learning Approach

Shalome Hanisha Tatapudi, Merck, Tampa, FL, United States

The decision-making process in healthcare, especially in the field of infectious diseases, is complex and challenging due to various factors such as disease transmission mechanisms, demographic variables, population behaviors, and available interventions. Disease modeling has proven to be an effective tool for evaluating the implications of policymakers' choices and quantifying the effects of decision-making.

However, there is still a need to navigate the trade-offs between cost-effective and efficient intervention strategies, dynamically. The evolving nature of disease and population further complicates the situation, as strategies that are effective today may become less effective in the future. Therefore, it is essential to examine how present intervention decisions will impact future outcomes and the health of the population at large. In this study, we propose the development of a reinforcement learning model to determine optimal intervention strategies, focusing on a case study of the human papilloma virus (HPV). By developing an intelligible learning model, we can evaluate comprehensive interventions, such as cervical cancer screening and vaccination strategies, at different time intervals and aim to create an optimal policy that spans an extended period. This research will help validate current decisions and provide insights into long-term impact of our choices on population's health. Designing effective intervention strategies is crucial for reducing the prevalence of HPV-related diseases and improving overall health outcomes. By leveraging accordingly, we can identify the most efficient and cost-effective approaches to HPV intervention and prevention.

3 - Analyzing BMI Dynamics During Pregnancy Using Growth Mixture Modeling and Machine Learning Techniques

Ashaar Rasheed, Binghamton University, Binghamton, NY, United States, Sreenath Chalil Madathil

Pregnancy induces various physiological changes, including Body Mass Index (BMI) fluctuations. Understanding these changes is crucial for monitoring maternal health. This study reveals different BMI trajectories across pregnancy using a growth mixture model (GMM), incorporating covariates such as age, race, and ethnicity to enhance interpretability. We analyzed BMI values at three time points—from the first to the third trimester—in a cohort of 234 pregnant women, employing 16 GMMs to identify distinct trajectories. Model selection included the use of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Integrated Completed Likelihood (ICL).

Our analysis identified three unique BMI trajectories, each characterized by average and variance parameters. The direct inclusion of age, race, and ethnicity in the GMM provided deeper insights into the variations in BMI trajectories. In addition to GMM, tree-based algorithms and regression analyses were used to explore the impact of additional covariates, including CDC morbidity indicators, diabetes, behavioral risk assessments, and delivery outcomes. This further enhanced our understanding of the factors influencing BMI changes.

The combined use of GMM and machine learning methods helped classify pregnancies into normal and abnormal categories, offering a nuanced view of body fat changes during pregnancy. This can inform health policies and practices aimed at improving maternal outcomes. The research emphasizes the importance of personalized health monitoring, which can lead to targeted interventions to optimize pregnancy outcomes.

Keywords: Pregnancy, BMI Dynamics, Growth Mixture Modeling, Covariate Analysis, Maternal Health

4 - A Theory-Based Framework for Understanding the Effectiveness of Generative Artificial Intelligence for Mental Support

Xunyu Chen, Virginia Commonwealth University, Richmond, VA, United States, Lingyu Li, Xiaojin Liu, Victoria Yoon

With its remarkable capabilities for language understanding and generation, Generative Artificial Intelligence (GAI), exemplified by ChatGPT, shows a high potential to address the global shortage of mental health support. However, little is known about how help-seekers perceive mental support provided by GAI, and consequently, how GAI can be improved to offer high-quality mental support. To address these questions, we propose an analytical framework grounded in communication and social support theories. Utilizing this framework, we develop multiple hypotheses regarding GAI's effectiveness in providing mental support and identify various moderators of its effectiveness, based on a comprehensive analysis of GAI's training and inference processes. The proposed hypotheses are tested using user-generated textual and evaluation data collected from a leading online mental healthcare community. Our findings not only reveal the unique attributes of GAI's mental support but, more importantly, provide systematic and actionable pathways to enhance GAI's effectiveness in delivering automated mental health support.

5 - Small Area Estimation of Case Growths for Timely Covid-19 Outbreak Detection

Zhaowei She, Singapore Management University, Singapore, Singapore, Zilong Wang, Jagpreet Chhatwal, Turgay Ayer

 Rapid and accurate detection of local outbreaks is critical to tackle resurgent waves of COVID-19. A fundamental challenge in case growth rate estimation, a key epidemiological parameter, is balancing the accuracy vs. speed tradeoff for small sample sizes of counties. We develop an algorithm, Transfer Learning Generalized Random Forest (TLGRF), that balances this tradeoff. Through transfer learning, TLGRF can accurately estimate case growth rates for counties with small sample sizes based on relevant day and county-level features affecting the disease spread. TLGRF outperforms established growth rate estimation methods and demonstrated that it can greatly improve the timely detection of outbreaks. We thus developed a open source tool for timely detection of COVID-19 outbreaks in each U.S. county, which received substantial attention by policymakers.

SC47

Summit - 439

Decision Making in Healthcare

Invited Session

Health Applications Society

Chair: Oguzhan Alagoz, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Yifan Lu, University of Wisconsin-Madison, Madison, 53706, United States

1 - A Transfer Reinforcement Learning Approach for Precision Follow-up Colonoscopy Recommendation for Colorectal Cancer Screening

Mu Du, School of Economics and Management, Dalian University of Technology, Dalian, China, People's Republic of, hongtao yu, Nan Kong

In precision public health (PPH), making well-informed intervention decisions tailored to sub-populations' disease progressions is challenging because of the scarcity of relevant data on target sub-populations. This study addresses precision follow-up colonoscopy recommendations for colorectal cancer prevention in multiple cohorts, whose medical data can only be obtained progressively during their follow-up visits. Specifically, each sub-population's disease progression is modeled as a Markov decision process (MDP) with unknown transition parameters and rewards based on a well-established microsimulation colon model. To solve the MDPs, we propose a learning-while-optimization framework with a transfer reinforcement learning (RL) approach facilitated by a general upper confidence bound value iteration scheme. We design this transfer RL approach by integrating a model-based RL with a joint estimator, which can combine huge proxy data historically collected from the entire population and scarce true data progressively obtained from target sub-populations. In addition, we accelerate the proposed transfer RL approach by modifying its algorithmic procedure to employ short-term planning for value iteration instead of full planning. We establish a theoretical performance guarantee by proving the regret bound of the proposed transfer RL approach. We also prove the accelerated transfer RL can significantly decrease the computational complexity without suffering performance degradation in regret. To the best of our knowledge, there is no regret guarantee for such transfer RL approaches, even in the tabular MDP settings. By integrating the RL techniques and operations research, our study pioneers an actionable plan for precision follow-up screening.

2 - Optimizing HIV Point-of-care Testing Systems in Kenya Under a Conditional Value at Risk Objective

Yinsheng Wang, University of Washington, Seattle, WA, United States, Shan Liu, Chaoyue Zhao

Minimizing the turnaround time for HIV testing by allocating point-of-care testing machines in a hub-and-spoke network of clinics has been proven effective in enhancing HIV care. Our previous work has successfully developed practical decision-support tools using queuing theory for HIV viral load and drug resistance testing in Kisumu County, Kenya. The data from Kisumu County between 2019 and 2023 reveal a distinct seasonal trend in viral load testing demand, with notable surges each spring. In light of the demand variations, this study extends a deterministic mixed integer programming model to a two-stage stochastic model. In this framework, the allocation of testing machines is determined in the first stage, while the referral network from clinics to hubs hosting these machines is optimized in the second stage. We have developed and solved a queueing-location-allocation model that incorporates a Conditional Value at Risk objective to manage uncertainties and risks associated with fluctuating testing demands. To ensure computational efficiency, the solution process employs Benders decomposition and heuristic cuts to expedite convergence. Computational results based on a HIV viral load testing network in Kisumu County are presented to compare the robust model to its deterministic counterpart and to demonstrate the robustness of the model.

3 - Early-Stage Clinical Trials with Patient Choice

Connor Van Ryn, Clemson University, Clemson, SC, United States, Amin Khademi, Qi Luo

Early-stage clinical trials (CTs) seek to identify safe dosages for new drugs for later trial phases. Clinical investigators (CIs) often face difficulty in the recruitment and retention of patients during these early stages, which can cause substantial expense and delay to the trial. The prevailing methods used in CTs are dose-escalation methods, where consecutive groups of patients receive increasing doses until significant toxicities are seen. Patient dosing is entirely structured and scheduled by the CI, with the patient having no input. This form of trial is highly conservative in its efficacy and toxicity and gives no agency to patients. This research involves the creation of a new CT design method which balances CT participant satisfaction and CI information gain by increasing the control patients have while still ensuring that adequate information about the drug is learned. The model being developed is a psuedo-market mechanism in which patients are attempting to maximize their expected health gain from different dosages, and CIs are attempting to maximize the information gained about efficiency and toxicity of the drug across different dosages.

4 - Personalized screening policy based on individual women's breast cancer risk

Yifan Lu, University of Wisconsin-Madison, Madison, WI, United States, Oguzhan Alagoz, Eugenio Quessep, Rick Groeneweg, John Hampton, Karla Kerlikowske, Harry Koning, Kathryn Lowry, Diana Miglioretti, Jeanne Mandelblatt , Clyde Schechter, Nicolien Ravesteyn, Brian Sprague, Natasha Stout, Anna Tosteson, Amy Trentham-Dietz

Most existing breast cancer screening guidelines focus only on average-risk women, which may lead to over-screening of low-risk women and under-screening of high-risk women. Several successful risk assessment models exist, which provide an individual woman's risk of developing breast cancer in the next several years (E.g. 5-year risk), however, their impact on long-term breast cancer outcomes is unknown. We utilized University of Wisconsin Breast Cancer Simulation Model, a member of National Cancer Institute-funded Cancer Intervention and Surveillance Modeling Network (CISNET) breast cancer working group, to incorporate risk-based screening based on risk assessment models. CISNET models have been used by policy makers in 2009, 2016 and 2024 to set the breast cancer screening guidelines in the US. We developed and evaluated several personalized screening strategies based on an individual's breast cancer risk to improve the balance of benefits and harms as well as better allocate the clinical screening resources more efficiently.

SC48

Summit - 440

Machine Learning Applications on Inventory Models

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Jingkai Huang, Zhejiang University, Hangzhou, N/A

Co-Chair: Kevin Shang, Duke University, Durham, NC, United States

1 - Data-Driven Optimal and Myopic Policies for Inventory Systems with Demand Covariates

Jingkai Huang, Zhejiang University, Hangzhou, China, People's Republic of, Kevin Shang, Yi Yang, Weihua Zhou

We study a multi-period backorder inventory system in which the random demand depends on exogenous covariates. The main goal is to develop a data-driven inventory policy that minimizes the total expected inventory cost. Based on the optimal policy structure under full distribution information, we first propose the Data-Driven Optimal (DDO) policy which utilizes a ``first predict, then optimize the residual-based empirical dynamic programming" framework, and provide its finite-sample performance bound. However, the DDO policy may suffer the curse of dimensionality, especially when the sample size is large. We then propose the simple myopic policy with covariates. With known distribution information, we characterize its optimality conditions and optimality ratio. With unknown demand distribution, we introduce the Data-Driven Myopic (DDM) policy and provide its finite-sample performance bound. Interestingly, we can show that the DDM policy performs the same as the DDO policy under a small sample size. We further propose another simple policy, called semi-myopic policy which takes the future costs into account, while maintaining a simple structure. The idea is to approximate cost-to-go functions by those from the classic multi-period inventory system without covariates which can be easily solved. We prove that the optimality condition of semi-myopic policy is looser than that of myopic policy. Accordingly, we also propose Data-Driven Semi-Myopic (DDSM) policy for the unknown demand case. Finally, we extend the main results to the scenario with correlated demand covariates. Numerical studies demonstrate that both heuristic policies perform well, while the semi-myopic policy performs better.

2 - Online Learning for Dual Index Policies in Dual Sourcing Systems

Jingwen Tang, University of Miami, Coral Gables, FL, United States, Boxiao Chen, Cong Shi

We consider a periodic-review dual-sourcing inventory system with a regular source (lower unit cost but longer lead time) and an expedited source (shorter lead time but higher unit cost), under carried-over supply and backlogged demand. Unlike existing literature, we assume that the firm does not have access to the demand distribution a priori and relies solely on past demand realizations. Even with complete information on the demand distribution, it is well-known in the literature that the optimal inventory replenishment policy is complex and state-dependent. Therefore, we focus our attention on a class of popular, easy-to-implement, and near-optimal heuristic policies called the dual-index policy. The performance measure is the regret, defined as the cost difference of any feasible learning algorithm against the full-information optimal dual-index policy. We develop a nonparametric online learning algorithm that admits a regret upper bound of O(\sqrt{T log T}), which matches the regret lower bound for any feasible learning algorithms up to a logarithmic factor. Our algorithm integrates stochastic bandits and sample average approximation techniques in an innovative way. As part of our regret analysis, we explicitly prove that the underlying Markov chain is ergodic and converges to its steady state exponentially fast via coupling arguments, which could be of independent interest. Our work provides practitioners with an easy-to-implement, robust, and provably-good online decision support system for managing a dual-sourcing inventory system.

3 - Dynamic Pricing with Infrequent Inventory Replenishments

Boxiao Chen, University of Illinois Chicago, Chicago, IL, United States, Menglong Li, David Simchi-Levi

We consider a joint pricing and inventory control problem where pricing can be adjusted more frequently than inventory ordering decisions. We consider the situation where the demand-price function and the distribution of random demand noise are both unknown to the retailer. We propose an online learning algorithm that achieves the optimal convergence rate to the true optimal solution.

4 - Multi-Period Newsvendor Optimization for Capacity Planning

Gah-Yi Ban, Imperial College Business School, London, United Kingdom, Abhilasha Katariya, Chinmoy Mohapatra, Liron Yedidsion

Capacity planning in Amazon Logistics is inherently a multi-period decision problem, whereby decisions are made and adjusted weeks in advance of deliveries. We formulate and solve a multi-period newsvendor optimization model for Amazon Logistics capacity planning. Since the general multi-period problem is computationally intractable, we propose four computationally efficient heuristic solutions: (i) a solution equivalent to the single-period newsvendor solution (NV), which is reflective of the current state-of-the-art, (ii) an extension of the newsvendor solution that takes ramp-up and down costs into account (NV+), (iii) a linear decision rule-based solution (LDR) and (iv) a hybrid of NV+ and LDR solution to solve the problem numerically. Our numerical results show that the multi-period solutions (LDR and Hybrid) outperform the single-stage solutions (NV and NV+) when the demand model exhibits non-stationarity and dependence over time, which can translate to substantial cost savings.

5 - Two-Sided Pricing and Learning with Inventory Constraints

Meichun Lin, Singapore Management University, Singapore, Singapore, Tim Huh

Motivated by online used-car platforms, we study pricing decisions for purchasing and selling a product in a two-sided market. With uncertainty from both supply and demand, a platform sequentially adjusts purchase and selling prices to maximize profit while satisfying inventory constraints. Moreover, the platform does not know in advance how supply and demand depend on the prices. Our work investigates how the platform can manage demand and supply in the presence of the two sides of uncertainty and inventory constraints. When the demand and supply functions are known, we show that simple fixed-price policies can be implemented with a small performance loss. Given limited information on supply and demand, we propose pricing algorithms that achieve the best possible regret when the planning horizon is large, which sheds light on how to balance the trade-off between two sides of learning and profit maximization (earning).

SC49

Summit - 441

New Advancements in Partially Observable Markov Decision Processes (POMDPs)

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Hao Zhang, University of British Columbia, Vancouver, BC, V6S 0H1, Canada

1 - Exact Solution to a Machine-Maintenance Problem with Multiple Unobservable States

Cong Yang, University of British Columbia, Vancouver, BC, Canada, Hao Zhang

We study a classic machine maintenance problem with multiple unobservable machine states. The state of the machine deteriorates according to a Markov process. The decision maker can choose from three actions---production, inspection, and replacement. Assuming perfect inspection, perfect replacement, and production without monitoring, the model is simplified to its essence while still encompassing a comprehensive and intricate optimal solution. We identify several possible structures of the optimal solution, which can be conveniently represented by graphs with multiple absorbing cycles. These cycles describe the steady-state phase of the optimal policy, which is endowed with desired features such as 'condition-based maintenance', 'dynamically scheduled inspections', and 'postponed replacement after an inspection'. Based on the analysis of these structures, we develop a policy iteration algorithm capable of finding the exact optimal solution in finite time. We also obtain a full characterization of the optimal solution in the three-state special case. This study fills a visible gap in our quest for optimal solutions for machine maintenance problems.

2 - Structural Estimation of Partially Observable Markov Decision Processes and Their Psychology Applications

Yanling Chang, Southern Methodist University, Dallas, TX, United States

Partially Observable Markov Decision Processes (POMDPs) is a well-developed framework for sequential decision making under uncertainty and partial information. This paper considers the (inverse) structural estimation of the primitives of a POMDP based upon data in the form of sequences of observables and implemented actions. We analyze the structural properties of an entropy regularized POMDP, specify conditions under which the model is identifiable without knowledge of the state dynamics, and develop a soft policy gradient algorithm to compute a maximum likelihood estimator. We apply the method to learn (i) mental fatigue dynamics which cannot be accurately observed in workplaces, and (ii) the net subjective task value for exerting cognitive control under different fatigue levels. Tested via a standard cognitively demanding n-back task, the method accurately inferred hidden fatigue dynamics and the agent's reward function. More importantly, the simulations and predictions of the fitted model are consistent with classic findings from existing empirical studies.

3 - Efficient Discovery of Cost-Effective Policies for Sequential Medical Decision-Making Problems

Narges Mohammadi, Imperial College London, London, United Kingdom, Reza Skandari, Anand Shah

Cost-effectiveness analysis (CEA) is extensively employed by healthcare policymakers to guide funding decisions and inform optimal design of medical interventions. In the CEA literature, willingness to pay (WTP) serves as a common metric for converting health benefits into monetary value and defining the net monetary benefit of an intervention. However, there is no universally accepted value for WTP. To address this issue, we propose presenting policymakers with a comprehensive menu of strategies that are proven cost-effective across a reasonable range of WTP values. In our approach, we consider a setting where the medical decision-making process can be formulated as a parametric linear programming model. We have developed a novel algorithm aimed at efficiently constructing the menu of cost-effective policies. Our algorithm is particularly suited for Constrained Markov Decision Process (CMDP) and Constrained Partially Observable Markov Decision Process (CPOMDP) models, which are commonly utilized modeling frameworks for addressing sequential medical decision-making problems. We have applied our modeling framework to design hearing loss screening strategies for cystic fibrosis patients. Informed by a validated, data-driven model, we have developed several heuristic and approximate policies, allowing policymakers to balance between performance and ease of implementation.

4 - Optimal Learning and Management of Threatened Species

Jue Wang, Smith School of Business, Queen's University, Kingston, ON, Canada, Xueze Song, Roozbeh Yousefi, Zhigang Jiang

Amid an unprecedented loss of biodiversity, a pressing issue is how to improve the efficiency of conservation with limited resources and information. Collecting data on species with a small population is costly and time-consuming, and many high-stake decisions need to be made based on limited data. We develop a partially observable Markov decision processes (POMDP) model with unknown parameters to jointly optimize the information collection and protection efforts for threatened species. The model takes into account uncertainties about the state, detectability, and dynamics of the species, and adaptively adjusts the efforts of surveying and protection in real time. Although the standard formulation is intractable, we exploit the structure of ecological problems to identify a hybrid belief state in low dimension, and reformulate the stochastic dynamic program as a piecewise deterministic optimal control problem. We also conduct a case study on the conservation of Hainan Gibbon, the rarest primate species, in which we extend the model to optimize the spatiotemporal allocation of limited resources.

SC50

Summit - 442

Retail Analytics

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Zumbul Atan, Eindhoven University of Technology-Industrial Engineering &Innovation Sciences, Eindhoven, Netherlands

Co-Chair: Alexander Huebner, TU Munich, Straubing, Germany

1 - Perishable Inventory Management under Freshness Dependent Stochastic Demand and Backroom Operations

Ilgin Efe Senyuva, Eindhoven University of Technology, Eindhoven, Netherlands, Melvin Drent, Zumbul Atan

We study a discrete-time perishable inventory management problem in which a retailer sells a single product whose lifetime decays linearly with time. Consumers value products with a higher remaining lifetime more than they value ones with lower remaining lifetimes. We assume that there is only a single batch of products with the same lifetime on the shelves. Periodically, the retailer can replace the old batch of products that are not yet sold with a new batch of fresh products. This induces a higher demand, but comes at a discarding cost. We formulate this problem as a Markov decision process and find the optimal policy through value iteration. Through simulations, we compare the optimal policy with myopic, near-myopic, base-stock and base-lifetime heuristic policies. The numerical experiments provide insights into the practical settings in which each heuristic performs well. Finally, we study multiple extensions to our problem. We include backroom operations in which products can be stored that decay at a lower rate. Here, we assume a non-fixed remaining lifetime when the replenishment is made, and make the replenishment batch size dependent on the displayed lifetime to account for products being mishandled at the backroom. We use the heuristics and optimal policies for our original problem as heuristic policies in these extensions and we show numerically how well they perform.

2 - Dynamic Cart Recommendation with Demand Learning

SHIWEI CHAI, University of Florida, Gainesville, FL, United States, Yining Wang, Lucy Zhang, Xiajun Pan

We consider an assortment recommendation problem in the context of an online shopping platform. Each customer, after arriving at the platform, picks up a primary product and adds it to the shopping cart. Based on the choice of the primary product, the platform then recommends at most K auxiliary products to the customer to maximize its revenue. We characterize customers' decision to buy an auxiliary product by a multinomial logit choice model, but the valuation of each product is unknown to the platform. This problem is closely related to the so-called MNL-bandit problem (Agrawal et al. 2019), but since the auxiliary product is recommended after observing the choice of the primary product, providing the same assortment repeatedly is not possible in our setting (which however is the key component in the algorithm design to derive unbiased estimators in Agrawal et al. (2019)). To overcome this challenge, we propose a new algorithm (MLE-UCB) based on the maximum likelihood estimation. Finally, a case study using the calibrated real-world data from a large online shopping platform in north America shows the proposed algorithm performs well empirically.

3 - Optimizing Assortment and Inventory on Shelves When Products Perish

Alexander Huebner, TU Munich, Straubing, Germany

Optimizing retail assortments is essential to maximize revenues, but also can have an impact on inventory ages and therefore product expiration. The limiting condition for assortment decisions is the shelf space of retailers. Assuming a constant available, but in size restricted shelf space, the assortment size also defines how much shelf space each product gets, and therefore how much items of each product can be stored on the shelf. That is why assortment planning is also important in terms of inventory management. Especially for perishable products, the inventory level on the shelf has to be carefully chosen in order to prevent deteriorating overstocks. Current literature considers single-period assortment and shelf space models, despite products can be stored and replenished over multiple sales periods, leading to different expiration dates on the shelf. We contribute with an assortment and shelf space model that considers replenishment and customer withdrawal over several periods, taking into account stochastic demand, space elasticity, substitution effects and customer's freshness preferences. To solve the model, we use an iterative heuristic approach with a recurring update of demand to account for substitution effects and their impacts on the assortment and shelf space decisions.

4 - Multi-product pricing: A customer choice model and a dynamic pricing approximation

Zumbul Atan, Eindhoven University of Technology-Industrial Engineering &Innovation Sciences, Eindhoven, Netherlands, Laura Sprenkels, Ivo Adan

We study a retailer's pricing problem of substitutable products. Demand is influenced by selling prices and availability of inventory. Assuming stationary demand process, we aim to find the optimal selling prices of the products. We study two practical variants of the problem. In the first one, we consider an infinite time horizon. Our model includes both customers' willingness to pay and substitution behavior explicitly, and can handle any type of correlation between products. For this problem, we present a new customer choice model. In the second variant, we consider a finite time horizon and finite amount of inventory, and we develop a dynamic pricing approximation for the selling prices.

SC51

Summit - 443

Sustainability and Technology

Invited Session

MSOM: iForm

Chair: Rowena Gan, Cox School of Business, Southern Methodist University, Dallas, TX, United States

Co-Chair: Rong Li, Syracuse University, Syracuse, NY, United States

1 - Contract Tokenization in the Renewable Energy Market

Rowena Gan, Southern Methodist University, Dallas, TX, United States

Endorsed by the blockchain technology, supply chain contracts can be digitally recorded and stored in crypto tokens, which is referred to as being tokenized. Tokenized contracts offer new ways of financing, trading and owning an asset. Using the renewable energy market as a backdrop, we study the impact of contract tokenization on different parties in the supply chain based on their respective incentives.

2 - Delivery Terms for Voluntary Carbon Offsets

Safak Yucel, Georgetown University, Washington, DC, United States, Vishal Agrawal, Gokce Esenduran

In addition to abating emissions, corporations purchase carbon offsets for voluntary decarbonization targets, such as reaching net-zero emissions. A carbon offset represents emissions reduction achieved by a developer through investments in projects, such as forestry and renewable energy. Corporations purchase offsets under two delivery terms. The first is forward delivery, where the buyer orders a certain quantity before the seller undertakes the investment and yield uncertainty realizes. The second is prompt delivery, where the seller invests and uncertainty realizes, before the buyer orders offsets. Motivated by the importance of choosing the right delivery term for buyers and sellers, in this paper, we investigate economic and environmental implications of these delivery terms. Our results offer several managerial insights: First, a seller should prefer forward delivery if investment cost is low, but prompt delivery otherwise. Although one would expect that the delivery term that enables the seller to make a higher investment results in a higher profit, we find that prompt delivery may lead to a higher profit despite lower investment. Second, a buyer should prefer forward delivery only for a project with either a low or high investment cost.

Interestingly, the delivery term that leads to a lower cost for the buyer may actually require higher abatement. Finally, environmental groups should promote forward delivery only for a project with a low investment cost. Moreover, when the buyer and seller prefer different delivery terms, the one preferred by the seller leads to greater emissions reduction, and should be promoted by environmental groups.

3 - Integrated Planning of Power System and Hydrogen Supply Chain

Siqiang Guo, University of Missouri-St. Louis, St. Louis, MO, United States, Erhan Kutanoglu, Shadi Goodarzi

Hydrogen is important as it can serve as a carrier of clean energy, providing decarbonization solutions to sectors that are highly carbon-intensive and hard to electrify (e.g., steelmaking). As the proportion of renewable energy in the power system continues to grow, hydrogen's role in mitigating the energy supply and demand discrepancies (both temporally and spatially) will become even more crucial. However, the hydrogen economy currently faces many challenges, such as the high cost of green hydrogen and the lack of adequate infrastructure. The diversity of the power system and hydrogen supply chain (in terms of energy/hydrogen production, transportation, and storage) and the close correlation between them complicate the expansion of both systems. This complexity is intensified by the potential influence of various environmental policies such as carbon allowance, carbon tax, and clean hydrogen subsidy. We provide a model for the joint optimization of the power system and hydrogen supply chain, and we delve into the aforementioned environmental policies, examining their impact on the entire system. Our detailed analysis reveals the rationales of different policies and yields several efficient policy combinations that can achieve significant carbon reduction (50 - 80%), maintain corporations' decision-making flexibility, ensure a low decarbonization cost for both corporations and society (< $65/ton of CO_2), while not adding significant burdens to corporations (< 3% cost increase).

4 - Collective Power Purchase Agreement Negotiation and Risk Sharing

Xiaoxuan Hou, University of Washington, Seattle, WA, United States, Shi Chen

Over 400 companies spanning more than 175 markets worldwide have committed to 100% renewable electricity. But achieving these goals can be challenging. In the past decade, long-term power purchase agreement (PPA) has gained popularity among large corporations like Google and Amazon (buyers) as the main vehicle to source supply from renewable energy developers (the seller). Under a PPA, the buyer commits to purchasing the renewable energy from the seller at a fixed price for a predetermined length of time (usually 10 to 15 years). Yet signing a PPA remains difficult for small-to-medium-sized corporate buyers, who may be too small to sponsor a full project, not have the expertise, or lack sufficient negotiation power on their own. As a result, collective PPAs emerged as a potential solution. We build Nash bargaining-based economic models to compare the performance of three negotiation mechanisms for these small-to-medium buyers to sign PPA contracts collectively: separate negotiation, agent-based negotiation, and joint negotiation. Our analysis reveals the impact of different mechanisms on the renewable project size in equilibrium and different players' preference. Assuming risk neutrality, our findings indicate that the three negotiation mechanisms lead to identical equilibrium project sizes, thus contributing equally to adding renewable capacity to the grid. Yet conflicts of preference for different negotiation mechanisms exists among parties, and rising number of buyers simplifies the consensus-building process for adopting joint negotiations. Further, buyers' perception of risks due to the uncertainty of the future market prices can harm the achievement of the sustainable goals.

SC52

Summit - 444

Enhancing Service Operations in Digital Platforms and Healthcare

Invited Session

MSOM: Service Operations

Chair: Chloe Glaeser, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States

1 - Can Employees&Rsquo; past Helping Behavior be Used to Improve Shift Scheduling? Evidence from ICU Nurses

Yixin Iris Wang, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Zhaohui Jiang, John Silberholz, Deena Costa, Michael Sjoding

Employees routinely make valuable contributions at work that are not part of their formal job description, such as helping a struggling coworker. However, the degree to which the past helping behavior of employees scheduled to a shift impacts that shift’s operational outcomes remains an under-explored question. Our empirical results indicate that shift-level past helping measures are predictive of patient length of stay (LOS), more so than the broadly studied notion of team familiarity. Counterfactual analysis shows that relatively small changes in shift composition can yield a significant reduction in total LOS. Overall, our study suggests the potential value of shift scheduling using data on past helping behaviors, which may have promise far beyond the selected application to ICU nursing.

2 - Predictive Modeling of Relapse Rates and Simulation for Improved Early Intervention in Opioid Treatment Centers

Umit Celik, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States, Bradley Staats

This study employs machine learning to predict relapse rates among opioid treatment patients, aiming to enhance early intervention strategies. Utilizing a dataset from over 200 U.S. opioid treatment facilities, we develop a predictive model using ensemble machine learning techniques, including random forests and logistic regression. This model identifies patients at high risk of relapse, facilitating the creation of personalized intervention strategies.

We apply simulation techniques to evaluate the causal effect of these targeted interventions in real-world scenarios, focusing on improving treatment outcomes and resource allocation. Preliminary results indicate that our predictive approach significantly improves the accuracy of risk assessments, enabling healthcare providers to implement more effective and efficient interventions.

This research demonstrates the potential of machine learning in transforming opioid treatment strategies, ultimately contributing to better patient care and enhanced operational efficiency in healthcare settings.

3 - Multihoming on on-Demand Ride-Hailing Platforms

Jaelynn Oh, The University of Utah, Salt Lake City, UT, United States, Kaitlin Daniels

We study the informational value of single homing on on-demand ride-hailing platforms.

4 - The Impact of Entry Fee Increase on Emergency Department Demand: a Territory-Wide Study in a Universal Healthcare System

Hyun Seok (Huck) Lee, Korea University Business School, Seoul, Korea, Republic of, Eric Park, Timothy Rainer

Emergency department (ED) overcrowding is a global health problem as it disrupts EDs' role as a public safety net. Several countries have increased the financial burden on access to ED care by increasing the ED entry fee in hopes of reducing patient traffic to mitigate ED overcrowding. We perform a territory-wide study of the universal public health system of Hong Kong SAR and empirically investigate the impact of the increase in the entry fee for the ED in June 2017 from HK$100 (US$13) to HK$180 (US$23) on the behavior of the visit of patients in the ED. In addition, we examine the heterogeneous effect between different types of patients.

SC53

Summit - 445

Additive Manufacturing Applications in Supply Chains

Invited Session

MSOM: Supply Chain

Chair: Ali Parlakturk, University of North Carolina Kenan-Flagler Business School, Chapel Hill, NC, 27599, United States

Co-Chair: Nagarajan Sethuraman, University of Kansas, Lawrence, KS, United States

1 - Lower Price OR Entry Barrier? Technical Service Design for Customers of Varying Expertise: An Application to 3D Printing Rental Business

Yue Zhang, Pennsylvania State University, University Park, PA, United States, Brian Tomlin

Customers with short-term, low-volume 3D printing (3DP) needs can rent a printer from a manufacturer or distributor (hereafter, OEM), or use a third-party printing service provider (hereafter, print farm). The customers vary in their 3DP expertise and will therefore differ in their cost of attaining any quality if using a rented printer, whereas print-farm technicians are experts in 3DP. Customers can choose to rent a printer and conduct 3DP on their own to attain qualities that maximize their payoff, or to delegate their needs to print-farm technicians who attain the desired qualities for them. To compete with print farms, OEMs can choose to offer technical service that augments the expertise of customers who rent printers. Technical service can be provided along with printer rentals in different formats - some OEMs bundle service with rentals whereas others offer optional service. Moreover, OEMs can adopt a standard approach to offer technical service in an identical way to all customers, or alternatively adopt a customized approach that caters the technical service offering to each customer's expertise. In this paper, we explore both approaches to study how (in terms of magnitude, format and price) should technical service be offered, and how customer expertise distribution influences the optimal offering and the OEM profit.

2 - Delegated Customization Using Retail 3D Printing

Ali Parlakturk, University of North Carolina Kenan-Flagler Business School, Chapel Hill, NC, United States

Retail customization with partially 3D printed products with traditionally manufactured cores has gained significant interest in recent years as a means of offering customized products. we propose a modeling framework that allows for partial customization using 3D printing while also capturing the impact of customer waiting time, quality issues, and inventory management.

3 - Minimizing the Number of Late Parts in Additive Manufacturing

Michael Stott, Texas A&M University, College Station, TX, United States, Jon Stauffer, Chelliah Sriskandarajah

Additive Manufacturing is a process by which three-dimensional components and products are made via the addition of material in a layer-by-layer fashion. This manufacturing technique has begun to see serious commercial usage given its advantages in creating very dense or complex geometries as well as highly customizable components. This paper studies the Nesting and Scheduling problem within the Additive Manufacturing context and shows the problem of minimizing the Number of Late Parts is strongly NP-hard even if the nesting of jobs is given. We also study and evaluate the performance of efficient heuristic algorithms to minimize this measure for given nesting of jobs, Heuristic algorithms are extended to study the problem with job nesting as well. An extensive computational study is contacted to evaluate the performance of these algorithms. These results are not currently shown in the literature. Moreover, all complexity results with nesting are obtained using minimal problem instances in the sense that parts are identical and differ only in their length but width and height of the parts are same. Consequently, these problem instances do not require any complex 2D or 3D packing algorithms for finding the nesting of jobs. In this way, the complexity results obtained here are unique and provide insights in to the complexity of nesting problems with minimal problem instances. We also study related scheduling problems under generalized due date problems context.

4 - Predictive Spare Parts Printing Using Sensor Information

Ipek Tanil, Eindhoven University of Technology, Eindhoven, Netherlands, Rob Basten, Collin Drent

We present our research on predictive spare parts printing on-site with sensor information to benefit from the additive manufacturing (AM) capability on time. AM enables the production of spare parts with an extremely short lead time, thereby enabling a faster supply solution than regular replenishments from the original equipment manufacturer. The former is however more expensive than the latter, which leads to a trade-off. We investigate this trade-off in the context of spare parts provisioning for a partially observable degrading system. Specifically, we consider a system with embedded sensors that relay classification data of critical components. We then propose a Bayesian procedure to infer from this data and the confusion matrices of the classifiers whether these critical components are nearing failure. Using this Bayesian procedure, a decision maker responsible for the uptime of the system needs to sequentially decide whether to predictively print spare parts or replenish spare parts from the regular supplier. To analyze this sequential decision problem, we formulate a partially observable Markov decision process. In an extensive numerical study, we investigate the trade-off between predictive printing and replenishing parts from the supplier. Notably, our numerical analysis gives answer to when it is optimal to print or order a part, and how this decision is impacted by various input parameters. The findings of this study allow us to quantify the benefit of using AM for printing spare parts with sensor information.

SC54

Summit - 446

Agriculture and Food Technology and Innovation

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Yanchong Zheng, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: El Ghali Zerhouni, MIT, Cambridge, United States

1 - Generative AI Can Harm Educational Outcomes

Alp Sungu, The Wharton School of the University of Pennsylvania, Philadelphia, PA, United States, Hamsa Bastani, Osbert Bastani, Haosen Ge, Rachel Mariman, Ozge Kabakci

Generative artificial intelligence (AI) is poised to revolutionize how humans work and has already shown promise in significantly improving productivity. However, a key question remains: how does generative AI affect learning, specifically how humans acquire new skills while performing tasks? This skill learning is critical for long-term productivity gains, especially in areas where generative AI is fallible and human experts must check its outputs. We studied the impact of generative AI, specifically OpenAI's GPT-4, on human learning in high school math classes. In a field experiment involving nearly a thousand students, we deployed and evaluated two GPT-based tutors: one mimicking a standard ChatGPT interface (GPT Base) and one with prompts designed to safeguard learning (GPT Tutor). These tutors comprised about 15% of the curriculum in three grades. Consistent with prior work, our results show that access to GPT-4 significantly improves performance (46% improvement for GPT Base and 131% for GPT Tutor). However, when access was subsequently removed, students performed worse than those who never had access (17% reduction for GPT Base). This indicates that GPT-4 can harm educational outcomes. These negative effects were largely mitigated by the safeguards in GPT Tutor. Our findings suggest that students use GPT-4 as a "crutch" during practice, leading to poorer performance on their own. To maintain long-term productivity, we must deploy generative AI cautiously to ensure humans continue to learn critical skills.

2 - Ai Inventory Co-Pilots for Data-Driven Perishable Inventory Management

Yu Nu, Cornell University, New York City, NY, United States, Meng Qi, Elena Belavina, Karan Girotra

This study explores using AI tools for reducing food waste in commercial food service establishments. We first examine the impact of building computer vision based systems to accurately measure and classify food waste, which provides high-fidelity transaction-level data on how food is wasted in a food service establishment. We quantify the value of deploying such systems, with early pilot trials indicating 30%~40% waste reduction. Moreover, we develop AI co-pilots that assist kitchen managers in managing inventory with a food waste reduction objective. These co-pilots could be prescriptive in providing directions on how to order, or simply watch out and detect when managers may be subject to biases and make mistakes in food production.

3 - Food Safety Risk Propagation Within Agricultural Supply Chains

Jane Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Retsef Levi, El Ghali Ahmed Zerhouni, Yanchong Zheng

Before reaching consumers, food products flow through multiple supply chain locations, including production, processing, and distribution points. At each location, these products can be exposed to adulterants such as environmental contaminants, pesticides, and food additives which can increase their food safety risk. Thus, understanding how food safety risk "spread" within the supply chain can inform the design of timely and resource-efficient regulatory strategies to protect consumers. Leveraging a comprehensive self-constructed database of 11.3 million food safety test records in China, this study builds a unique panel data set to evaluate food safety risks propagation within the supply chain. Specifically, it develops a causal inference assessment of the product-level and adulterant-specific propagation effect of food safety risks from upstream wholesale markets and wet markets (WSMs/WMs) to more downstream supply chain locations such as retailers and restaurants. The key findings of this study highlight several regulatory insights. Notably, effective food safety risk management for perishable products can benefit from more frequent testing in WSMs/WMs. In particular, frequent testing for pesticides and veterinary drugs in WSMs/WMs can effectively prevent risk associated with these adulterants from propagating downstream. On the other hand, food additives/illegal non-food substances and environmental/microbial contaminants require additional inspection at the manufacturer and retail level.

4 - Financial Access or Price Premiums? A Nuanced View into Improving Farmer Welfare and Reducing Child Labor in Commodity Supply Chains

Andreas Kilian Gernert, KLU, Hamburg, Germany, Andre Calmon, Dan Iancu, Luk Van Wassenhove

Millions of smallholder farming households engaged in the production of agricultural commodities grapple with poverty and resort to their children's labor to boost production and enhance their living conditions. In response, governments, commodity-buying firms, micro-finance institutions, and non-profits have initiated various interventions to improve household welfare and mitigate child labor, but the effectiveness of these interventions remains uncertain. Drawing inspiration from cocoa supply chains, we propose a model that captures a farming household's need to secure basic subsistence, alongside its decisions on borrowing, saving, consumption, and the reliance on child labor for farm production. With this model, we evaluate the effects of common financial interventions, including improved access to credit and savings and the implementation of price premiums, examining their impacts on household welfare, immediate consumption, and child labor usage. Our analysis reveals a complex landscape: improved access to credit can either decrease or inadvertently increase child labor, depending on whether households borrow for subsistence or discretionary expenses.

Conversely, enhancing access to savings invariably curtails child labor but might simultaneously depress household consumption. Furthermore, price premiums can effectively diminish child labor usage only when they are significantly large; otherwise, they may inadvertently incentivize increased child labor. These results highlight the dual-edged nature of financial interventions and underscore the critical need for tailoring strategies to the specific operational dynamics of farming households, ensuring interventions do not inadvertently exacerbate the issues they aim to resolve.

SC55

Summit - 447

Civic Engagement and Community Analytics

Invited Session

Public Sector OR

Chair: EunSu Lee, New Jersey City University, Jersey City, NJ, 07311, United States

1 - Community Engaged Learning as a Tool for Learning Community Systems

EunSu Lee, New Jersey City University, Jersey City, NJ, United States

This research explains the practices of civic engagement and discusses directions to improve them in the classroom. During the students' activities, the concept of Public Operations Research was introduced, and students identified community issues and provided data-based decision making.

2 - "Exploring Digital Logistics Living Lab: Implementing Vehicle-to-Vehicle Transshipment for Small Business Expedited Delivery

HanByul Ryu, Inha University, Michuhol-gu, Korea, Republic of, Daisik Nam

With the growth of the e-commerce and logistics markets, the demand for expedited delivery services, such as same-day delivery, has been increasing. Consumers now expect same-day or next-day delivery as a standard, raising the expectations for logistics services, in turn, small businesses lose their competency with massive ecommerce platforms such as Amazon and Coupang. While previous studies have focused on the increased utility for consumers and the efficiency of carriers for same-day delivery, there is a lack of research on the utility of same-day delivery for small businesses. Small businesses, characterized by lower volumes and minimal inventory, require logistics services beyond traditional solutions like Fulfillment Centers, which are often designed with higher demand in mind. This study presents the planning and implementation of a digital logistics living lab project in collaboration with local governments in South Korea, driven by governmental initiatives and public-private partnerships. We demonstrate the viability of a logistics network utilizing vehicle-to-vehicle transshipment to achieve same-day delivery goals. We also analyze the impact of expedited delivery goals for small businesses and the community. Results indicate an increase in sales and social utility for small businesses. This research is expected to provide critical insights for the formulation of policies and strategies to support small businesses.

3 - Retirement decision-making: When is the right time?

Ilbin Lee, University of Alberta, Edmonton, AB, Canada, Mehrnaz Behrooz, Ivor Cribben

Retirement is a crucial stage in life and a significant life transition that requires careful planning and decision-making. Commencing retirement planning at younger ages enables individuals to make smarter decisions and provides them with a broader spectrum of choices as they age. In this research, we developed a Markov Decision Process (MDP) model to optimize retirement decisions while refining its parameters through data analytics of longitudinal datasets. Initially, we discerned which and how different factors, such as health, marital status, gender, income, etc., affect retirement determinants, including retirement age and type (voluntary/involuntary). Subsequently, we devised a personalized MDP model to delineate the optimal retirement age and savings strategy. Our model determined the optimal age for retirement based on factors such as health status, risk tolerance, income, and savings levels. Furthermore, the empirical results of our model suggest that retirement can be expedited by beginning to save at younger ages. These insights hold significance for policymakers aiming to refine pension programs and individuals striving to enhance their retirement prospects through informed planning.

4 - Conjoint Analysis and IPA of OTT service

Junho Lee, Korea University Business School, Seoul, Korea, Republic of, Hosun Rhim

We conduct text mining to extract attributes reflecting user opinions by collecting reviews of OTT users through web crawling. Conjoint analysis is performed to obtain part-worth for all attributes. IPA is used to identify attributes to prioritize sequence of the development.

5 - Conjoint Analysis of M-Commerce

Jaejun Hwang, Korea University Business School, Seoul, Korea, Republic of, Hosun Rhim

We develop a service strategy of M-commerce firms based on conjoint analysis. Attributes are identified by text mining. We used 7,820 Amazon, Alibaba review data.

SC56

Summit - 448

Implementing DEI Initiatives in Universities: Success Stories and Remaining Hurdles

Panel Session

Minority Issues Forum

Co-Chair: Veronica Villena, Arizona State University, Tempe, AZ, United States

Co-Chair: Jason Acimovic, Penn State University, State College, PA, United States

1 - Panelist

2 - Panelist

Jaime Campbell, Pennsylvania State University, University Park, PA, United States

3 - Panelist

Jeffrey Wilson, Arizona State University, Tempe, AZ, United States

4 - Panelist

Candace Yano, UC Berkeley, Berkeley, CA, United States

5 - Moderator Panelist

6 - Panelist

Rohit Verma, University of South Carolina, Ithaca, NY, United States

SC57

Summit - Terrace Suite 1

Operations Research and Machine Learning for Maternal Health

Invited Session

Health Applications Society

Chair: Meghan Meredith, Georgia Institute of Technology, Atlanta, United States

Co-Chair: Lauren Steimle, Georgia Tech ISyE, Atlanta, GA, United States

1 - A Stochastic Programming Approach for Patient Acceptance Decision-Making in Prenatal Care

Leena Ghrayeb, University of Michigan, Ann Arbor, MI, United States, Amy Cohn, Ruiwei Jiang, Alex Peahl

In current practice, depending on physician and clinic-related capacity constraints, prenatal care clinics define the number of patients they can accept by the number of patients that each physician can care for at any point in time. Once a patient is accepted to the clinic, they are considered under the provider's care for the entirety of their pregnancy. While this method is simple and provides clear guidelines, it does not account for the stochastic nature of patients' health states – patients may require additional appointments beyond their defined pathways due to complications during pregnancy. This can result in unexpected overutilization, overbooking, and poor scheduling practices, which are burdensome for clinics and can hinder access to care for patients. We propose a multi-stage stochastic programming approach to aid clinics in deciding how many patients to accept on a weekly basis, given patient-related randomness inherent in pregnancy. We derive upper and lower bounds for this model and draw insights about patient acceptance policies.

2 - Identify Medically Unnecessary Cesarean Deliveries

Emily Fainman, Texas State University, San Marcos, TX, United States, Beste Kucukyazici, TING WU

We first use semi-supervised learning methods to classify pregnancy women into low- and high- risk groups based on approximately 18 million individual birth records in the U.S. from 2016 to 2021.
We develop an innovative semi-supervising fuzzy clustering (S-FCMd) algorithm by introducing a distance measure for mixed data consisting of numerical and categorical variables. We develop an innovative semi-supervising fuzzy clustering (S-FCMd) algorithm by introducing a distance measure for mixed data consisting of numerical and categorical variables. The labelled observations incorporate prior medical knowledge and effectively speed up the convergence. The method of updating cluster medoid ensures a computational complexity proportional to the sample size. Our experiments on real-life and synthetic data demonstrate the efficiency of our algorithm for large datasets. Then, we validate the optimal delivery methods for two risk groups through post-delivery variables in the dataset. We perform statistical inferences to compare actual post-delivery complications resulting from natural births and C-Sections in each cluster, and thus figure out the optimal delivery mode for each patient group.

3 - The Role of Hospital Occupancy on Maternal Health Disparities

Rachna Shah, University of Minnesota, Minneapolis, MN, United States, Alison Murphy

Racial disparities in maternal health are well documented, with Black women 2.5 times more likely than non-Hispanic white women to experience a major complication or death. Unfortunately, both the overall rate of negative outcomes and the racial disparities are increasing. Social determinants of health (e.g., income, housing, and access to medical care) and structural racism are significant contributors to racial disparities. For example, perceived racism among Black women is associated with pre-term birth and racial disparities in infant health are higher in counties with higher implicit or explicit racial bias. In this study, we (1) Determine if the effects of provider workload are moderated by patient race. We ask if Black women are more impacted by workload than white women in the context of maternal health; (2) Determine if the above relationship is stronger in areas which higher level of implicit or explicit bias compared to areas with lower bias, and (3) Demonstrate the mechanism that we theorize to link workload with racial health disparities.

4 - Preventing Rural Obstetric Hospital Closures: a Multi-Level Optimization Approach

Meghan Meredith, Georgia Institute of Technology, Atlanta, GA, United States

Rural residents experience high rates of maternal mortality and travel far distances for obstetric care as rural obstetric hospitals are closing at an increasing rate. Rural hospital administrators are faced with challenging decisions to close their hospitals as they confront patient safety and financial viability concerns. In response to these closures, the US government has initiated many programs to prevent rural hospital closures and maintain access to high-quality obstetric care. We propose a multi-level optimization modeling approach to determine the optimal allocation of rural hospital closure prevention resources to maximize access to obstetric care. We incorporate multiple aspects of access, including quality and risk-appropriateness. A case study of the state of Georgia is presented to illustrate geographic access to obstetric care and the impact of optimally allocating resources.

SC58

Summit - Terrace Suite 2

Analytics for Social Good: Public Sector Operations

Invited Session

Health Applications Society

Chair: Xiaoquan Gao, Purdue University, West Lafayette, IN, United States

1 - Achieving Rawlsian Justice in Food Rescue

Gerdus Benade, Boston University, Boston, MA, United States, Aydin Alptekinoglu

We explore a practical idea to achieve more fair outcomes in food rescue platforms that operate on a first-come-first-serve basis: Give priority to a select set of recipients – first dibs in claiming a food item – for a limited time window. We develop a general model of priority lists and show several structural results that characterize the optimal priority list for a given donation. The objective is to maximize the benefit (defined as pounds of food received) gained by the worst-off recipient on the platform, which is a Rawlsian notion of fairness. The model captures priority lists in their most general form: For each donation, it sets a release time for each potential recipient after which they will be notified of the donation. It can also accommodate perishability of food donations. Intuitively, our structural results show that it's optimal (in a max-min sense) to give higher priority to recipients that have received less than others in the past (so, they are currently among the worst-off) and to recipients that are slower than others in responding to posted donations. A simple index combines these two characteristics of recipients and provides the optimal priority ordering. We then construct an efficient algorithm to compute the optimal priority list for a given donation, which (in its most general form) boils down to setting notification times for each recipient. Finally, we present counterfactuals that quantify the potential impact of implementing this algorithm in practice by using historical data from our partner platform's Florida operations.

2 - Volunteer Management

Xinyuan Zhang, UNIVERSITY OF NOTRE DAME, South Bend, IN, United States, Eunae Yoo, Alfonso Pedraza-Martinez

Online volunteering platforms allow individuals to contribute to humanitarian projects remotely. This paper investigates user behavior in online volunteering platforms integrated within larger online communities with broader purposes. We study volunteers shifting between humanitarian and non-humanitarian projects, and its implications on volunteer retention.

3 - Enhancing Support for Survivors of Abuse Through Intelligent Volunteer Management

Rachel Wong, University of Toronto, Toronto, ON, Canada, Sheng Liu, Timothy Chan

In Canada, at least 1 in 5 women experience some form of abuse from a spousal partner with the economic repercussions of intimate partner violence (IPV) reaching an annual sum of $7.2 billion. Survivors risk losing everything they own when leaving an abusive relationship since moving and storing one's belongings is challenging due to financial, legal, and logistical barriers While shelters provide a temporary refuge, they seldom can assist with retrieval and/or storage of survivor's belongings. Shelter Movers stands out as the sole organization in Canada addressing this vital need. The organization uniquely offers free, survivor-focused moving and storage services, bridging a crucial service gap for those in need. Like many other non-profit organizations, they call on volunteers to provide essential services. As the demand surges, Shelter Movers face growing challenges in engaging and retaining volunteers to ensure a fast turnaround time for urgent service requests. Notably, fewer than five percent of the active volunteers in the pool regularly participate in move requests. The high volatility of service demand and volunteer availability further complicates the retention challenge. This work proposes an optimal model to match volunteers to 'move teams' considering underlying volunteer behavior based on past experiences. We utilize this matching and behavioral model in a multiclass queuing model to better understand system dynamics and make managerial recommendations for how Shelter Movers can better engage volunteers and improve volunteer retention.

4 - Keep Water Flowing: The Hidden Crisis of Rural Water Management

Chengcheng Zhai, University of Notre Dame, Notre Dame, IN, United States, Rodney Parker, Kurt Bretthauer, Alfonso Pedraza-Martinez, Jorge Mejia

In rural areas of sub-Saharan Africa (SSA), people rely on communal handpumps for clean drinking water. But these handpumps break down frequently. Thus, it is crucial to proactively maintain and reactively repair these handpumps to ensure continuous access to water. In this research, we study the optimal design of a water point maintenance program implemented by NGOs. We first conducted field research in Ethiopia and Malawi to better understand the context of water point maintenance. We then collected 56,344 water point functionality observations from NGOs implementing water point maintenance programs in the Central African Republic, Ethiopia, and Malawi. Lastly, we develop a Markov decision process that determines the optimal schedule for NGO mechanics to visit water points. We apply the optimization model to data from the three countries to identify the gap between practice and optimality, while exercising two heuristic policies (cyclic and responsive) observed in practice.

SC59

Summit - Ballroom 1

Advanced Modeling Techniques in Infectious Disease Management

Invited Session

Health Applications Society

Chair: Pinar Keskinocak, ISyE Georgia Tech, Atlanta, GA, United States

Co-Chair: Daniel Kim, Harvard Medical School & Massachusetts General Hospital, Boston, MA, United States

1 - The Benefits of Using Geospatially Detailed Agent-Based Models for Modeling Mitigation Strategies for Infectious Disease

Mark Roberts, University of Pittsburgh, Pittsburgh, PA, United States

The Public Health Dynamics Laboratory (PHDL) at the University of Pittsburgh has developed flexible, open-source agent-based modeling platforms to represent infectious diseases and evaluate prevention, mitigation, and treatment strategies across a wide variety of diseases, including influenza, Covid-19, measles, and others. By combining synthetic populations that are statistically equivalent to the real population in given areas, the models can incorporate characteristics of location and individual in their representation of the disease. This talk will describe successful representations of measles, influenza, and COVID-19.

We developed a national tool to example county-specific vaccination rates against measles after the Disneyland measles outbreak in 2014. Using location-specific data on schools, households, neighborhoods and workplaces, we could provide visual representations of the effect of decreasing vaccination for any county in the US. Similarly, using a modified SEIR model of influenza that operates within each individual in a simulation, we have investigated multiple vaccination and treatment strategies for both annual epidemic and pandemic influenza. Finally, by modifying the influenza model to represent the infectious and disease characteristics of COVID-19, we were able to estimate the impact of prevention and mitigation strategies in Pennsylvania prior to the widespread introduction of the COVID vaccine for the PA department of Health.

Agent-based models are particularly well suited to represent infectious diseases through their ability to alter the contact rates and infectivity of different people in different geographic locations.

2 - The Burden of Tuberculosis Among Foreign-Born Canadians: Estimates with Dynamic Models

Jeremy Chiu, Simon Fraser University, Burnaby, BC, Canada, Alexander Rutherford

Tuberculosis (TB) disease caused 1.6 million deaths worldwide in 2021. Foreign-born Canadians bear 80% of the burden of tuberculosis disease despite only comprising 23% of the total population in Canada. In this talk, we describe the pathology of TB and its unique characteristics that make it different from other infectious agents. We then illustrate how we investigate the impact immigration has on TB incidence.

We estimate the prevalence of latent TB infection (LTBI) among new immigrants by synthesizing census data with WHO reported tuberculin skin tests and smear positive tests. We also extend an SEIR (susceptible-exposed-infected-recovered) compartment model. The solution to the system of ordinary differential equations gives the distribution of actively infected, latently infected, and uninfected individuals among the foreign-born population. We then use Matlab's optimizer *fmincon*, a non-linear constrained optimizer, to infer unknown parameters by minimizing the least-squares differences between our model's computed incidence and prevalence vs Canadian reports.

We estimate that in 2014, 21.8% of new immigrants had latent TB infection, and that 0.73% were recently infected (2 years). Including this data in our model gives estimates that are consistent with Canadian reports of incidence and prevalence of active TB, a case study's findings on recurrent TB cases, and other estimates of LTBI-prevalence among foreign-born Canadians. This model can help inform policy to reach the Public Health Agency of Canada's goal to reduce incidence to <1 per 100,000 by 2035.

3 - A Multi-Agent Reinforcement Learning Framework for Evaluating the U.S. 'Ending the Hiv Epidemic' Plan

Dinesh Sharma, University of South Florida, Tampa, FL, United States, Ankit Shah, Chaitra Gopalappa

HIV remains a pressing public health issue in the United States, affecting approximately 1.2 million individuals, with 35,000 new infections annually. Disparities in HIV prevalence and access to care persist across different regions of the country. The 'Ending the HIV Epidemic (EHE)' initiative, launched in 2019, seeks to reduce new infections by 90% by 2030 by enhancing diagnoses, treatment, and prevention efforts, particularly in high-prevalence areas. However, existing decision models often focus solely on specific cities or the national population, failing to account for jurisdictional nuances and interactions. To address this gap, we propose a novel approach using multi-agent reinforcement learning (MARL) to analyze jurisdiction-specific interventions while considering cross-jurisdictional epidemiological dynamics. Our experimental analyses, conducted in California and Florida, reveal significant disparities between optimal policies derived from MARL versus single-agent reinforcement learning (SARL), underscoring the importance of accounting for jurisdictional variations and interactions. By leveraging comprehensive HIV modeling and tailored formulation of state space, action space, and reward functions, our study showcases the effectiveness of MARL in guiding public health policies and provides a scalable framework for informing the EHE at the national level.

4 - A Predictor-Corrector Algorithm for Estimating Time-Dependent Parameters of COVID-19 Models

Susan Rogowski, North Carolina State University, Raleigh, NC, United States

Stable parameter estimation is an ongoing challenge within biomathematics, especially in epidemiology. Oftentimes epidemiological models are composed of large numbers of equations and parameters. Due to high dimensionality, classic parameter estimation approaches, such as least square fitting, are computationally expensive. Additionally, the presence of observational noise and reporting errors that accompany real-time data can make these parameter estimation problems ill-posed and unstable. The recent COVID-19 pandemic highlighted the need for efficient parameter estimation tools. In this talk, we present a modified version of a regularized predictor–corrector algorithm aimed at stable low-cost reconstruction of infectious disease parameters. This method is applied to a novel compartmental disease model describing COVID-19 dynamics, which accounts for vaccination and immunity loss (from vaccinated and recovered populations). Numerical simulations are carried out with synthetic and real data for COVID-19 pandemic. We apply the predictor-corrector algorithm to reconstruct the time-dependent transmission rate and the effective reproduction number. This model and algorithm can easily be adjusted to be applied to other seasonal outbreaks.

SC60

Summit - Ballroom 2

Transportation and Network Optimization at Amazon

Invited Session

The Practice Section of INFORMS

Chair: Nicholas Kullman, Amazon, Bellevue, WA, United States

Co-Chair: Meltem Ozmadenci, Amazon, Seattle, WA, United States

1 - Optimizing Amazon's Network Flow Plan: A Decomposition Approach in Balancing Truck Fill Rates and Labor Utilization

Hyemin Jeon, Amazon.com, Bellevue, WA, United States, Theodoros Pantelidis, Martin Bagaram

Amazon's transportation network consists of different components whose operational plan rely on multiple layers of analytical tools to forecast, simulate, and optimize. As obtaining a solution that satisfies all business objectives concurrently is unrealistic over a large complex system such as Amazon's, many decisions are made isolated without the full awareness of other components. In this work, we introduce a coherent volume plan in order to maximize utilization of outbound truck capacity and labor supply on downstream processing sites concurrently. However, developing a mathematical formulation that captures all necessary dynamics between origin-destination pairs, i.e. fulfillment centers, sort centers and delivery stations has proven to be computationally intractable. Thus, we seek to decompose this complex problem into two components, one which contains the fulfillment center outbound network subgraph and aims to maximize truck fill rate and the other one that maximizes sort centers and delivery stations labor utilization. In this manner, each problem solves for a subgraph of the network and objective while remaining in parity via the propagation of minimum volume targets on fulfillment center outbound trucks generated by the second optimization problem that restrict the first problem solution. Hence, we approximate the global optimization solution for the entire network.

2 - COPPER: Contract Price Optimization Using Willingness-to-Pay Models for Amazon Freight

Yebin Tao, Amazon, San Mateo, CA, United States, Juan Xu, Roger Lederman

We develop COPPER, contract price optimization using execution and award models for willingness-to-pay (WTP), to replace the current manual and rule-based pricing mechanism for Amazon Freight (AF) contract. To capture shippers' WTP, we first build an award model to predict the probability of awarding a contract at a certain price, and then build an execution model to predict the number of executed loads given a contract awarded at a certain price. To estimate the optimal prices, we first consider each lane separately using simple grid search over a list of candidate prices, denoted as single-lane COPPER. As an enhancement, we further explore portfolio optimization by jointly considering lanes within the same portfolio (e.g., same pickup location), denoted as portfolio COPPER. We utilize coordinate descent to solve the high-dimensional portfolio price optimization efficiently. We conduct comprehensive model comparisons to build the WTP models and test the price optimization using the AF contracts that we bid on in 2023. From the test, single-lane COPPER shows significant improvements over the historical manual pricing with an estimated revenue increase of 13%, while portfolio COPPER would further increase revenue by another 10%.

3 - Air Cargo Revenue Management: A Two-Stage Stochastic Program with Learning-based Recourse

Yan Zhang, Amazon, Seattle, WA, United States, Nilay Noyan, Roger Lederman

In a logistics system, the space in trucks or airplanes unfilled by existing loads is considered as opportunity loss. To mitigate such loss, it is beneficial to find additional shippers who are willing to pay to use these space. However, finding such shippers could incur risk of displacing existing loads, due to the uncertain knowledge of the amount of unfilled space in the network. When displacement happens, existing loads are re-routed by a subsequent system at different displacement costs. To optimize the total revenue realized by the logistic system, it is critical to predict such displacement cost rendered by the subsequent routing system so that we can price shippers properly. To solve this problem, in this talk, we present a model-based machine learning approach to learn the routing decisions of the black-box system. Then, we incorporate the learned module into a two-stage stochastic programming problem, and solve the formulated problem using decomposition framework. The model estimates the value of the unfilled space in the logistic network, which will be used to set the price for shippers and to guardrail the revenue of the network.

4 - Accelerating Deliveries: Decomposition-based Approaches for Large-Scale Air Network Optimization

Tulio Toffolo, Amazon, Bellevue, WA, United States, Haroldo Santos, Valentina Vaca, Ruilin Ouyang, Wendian Wan, Na An

The Amazon Air Network Design (A2ND) tool is a patented model that optimizes flight schedule and package flow problems for Amazon Air. In recent years, features have been added to the model to incorporate optimization for package delivery speed, while maximizing aircraft utilization and minimizing operating cost. Over 70 types of operational constraints and a configurable objective function are used to explore network designs and perform trade-off analyses. As Amazon Air grew so did the size and complexity of the optimization model and computational performance became a bottleneck. This talk discusses the decomposition-based approaches developed to improve the scalability of the solver and the quality of the network solutions, including model generation and dynamic flight-set population. Computational experiments show a runtime reduction of more than 50%.

SC61

Summit - Ballroom 3

Navigating Impactful Careers Outside Academia: Insights and Perspectives

Panel Session

The Practice Section of INFORMS

Co-Chair: Sharon Arroyo, The Boeing Company, Sammamish, WA, United States

1 - Panelist

Sharon Arroyo, The Boeing Company, Seattle, WA, United States

Interested in exploring career paths beyond academia? Join us for an informative session where industry professionals will provide valuable insights and firsthand experiences about diverse and rewarding career opportunities. Our panelists will offer perspectives on their roles in various non-academic sectors, shedding light on the exciting and impactful work they do outside the traditional academic realm.

2 - Panelist

Rajeev Namboothiri, GE Vernova Advanced Research, Bangalore, India

3 - Panelist

Wotao Yin, DAMO Academy, Alibaba US, Bellevue, WA, United States

4 - Panelist

Genetha Gray, Edward Jones, Elk Grove, CA, United States

5 - Panelist

Gwen Spencer, Netflix, Seattle, WA, United States

6 - Panelist

Tim Jacobs, Amazon, Tempe, AZ, United States

SC62

Summit - Signature Room

Nonconvex, Nonsmooth, and Nonregular Optimization: A Computational Framework

Invited Session

TutORial

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Nonconvex, Nonsmooth, and Nonregular Optimization: A Computational Framework

Michael Ferris, University of Wisconsin, Madison, WI, United States, Olivier Huber, Johannes Royset

Algebraic modeling languages presently lack the ability to effectively support the formulation and solution of nonconvex and nonsmooth optimization problems. Since an arbitrary problem of this kind is intractable, any hope to achieve practically useful solutions would rely on means to convey specific problem structure to an algorithm. In this tutorial, we present a framework for specifying nonconvex, nonsmooth, and nonregular problems within an algebraic modeling language that makes available the key structural properties to an algorithm. It also facilitates experimentation with different model formulations and algorithmic approaches. The framework entails a change of mindset away from the traditional formulation of objective and constraint functions, and instead asks the analyst to specify a basic feasible set, a basic objective function, one or more monitoring functions, and several performance functions. Eleven examples ranging from goal programming to variational inequalities and engineering risk analysis illustrate the practical implications of the framework.

SC63

Regency - 601

Innovative Application of AI in Information Systems Research

Invited Session

Information Systems

Chair: Ali Tosyali, Rochester Institute of Technology, Rochester, NY, United States

Co-Chair: Jeongsub Choi, West Virginia University, Morgantown, WV, United States

1 - An Explainable Ai Framework for Identifying Deceptive Reviewers in E-Commerce Platforms

Ali Tosyali, Rochester Institute of Technology, Rochester, NY, United States

This study addresses the pressing issue of fake reviews on e-commerce platforms, which significantly impact consumer choices and threaten the credibility of online marketplaces. Recognizing the limitations of current methods focused on detecting individual fraudulent reviews, our research introduces a novel approach by examining the broader network of review activities. Leveraging insights from recent findings, we aim to identify and analyze the behavior of reviewers involved in fake review schemes. By using an advanced AI framework that integrates textual analysis, metadata, and network structure, we explore the characteristics that distinguish genuine reviewers from fraudulent ones. This comprehensive methodology not only enhances our ability to detect deceptive practices but also sheds light on the underlying mechanisms of online review manipulation, offering new avenues for improving the transparency and trustworthiness of online review ecosystems.

2 - Xscanner An Explainable Ai Approach for Comprehensive Screening and Pattern Analysis

Salih Tutun, Washington University in St Louis, St Louis, MO, United States, Ali Tosyali, Kazim Topuz, Anol Bhattacherjee

In this research, we introduce XScanner, an innovative explainable AI (XAI) approach designed for comprehensive screening and pattern analysis across various problems. XScanner leverages clinical data and protocols to ensure reliability, employing an XAI framework to predict outcomes and provide clear explanations of the underlying features influencing each prediction. This transparency is critical for building trust and acceptance of AI applications. XScanner distinguishes itself by generating full-color diagnostic images that visualize patterns in the data, akin to radiological images, offering a novel way for experts to interpret AI-driven insights. Unlike traditional black-box AI models, XScanner's explainability enhances its utility by making the decision-making process accessible and understandable to experts, thereby improving outcomes through informed decision-making. Our approach has been rigorously evaluated through various studies, demonstrating high classification accuracy with average F1 scores. These results underscore XScanner's potential as a powerful tool for early screening and analysis, capable of addressing the critical gaps in different business problems. XScanner represents a significant advancement in the screening and analysis. This research highlights the importance of transparency and explainability in AI, paving the way for future innovations that can enhance the accessibility and effectiveness of mental health care.

3 - Adaptive Sparse Pca: Enhancing Interpretability and Robustness in High-Dimensional Data

Yifan Xie, Rutgers University, Piscataway, NJ, United States

The surge in high-dimensional datasets has highlighted the limitations of traditional Principal Component Analysis (PCA), particularly the curse of dimensionality and the need for interpretable, sparse representations. Sparse PCA has emerged as a valuable tool to address these issues, but existing methods often rely on fixed thresholds, which can be suboptimal and inflexible. We introduce Adaptive Sparse PCA (AS-PCA), an innovative method that dynamically adjusts the sparsity threshold based on the data's intrinsic characteristics at each iteration. This approach ensures better adaptability, robustness to noise, and computational efficiency. By focusing on sparse principal components, AS-PCA enhances interpretability and generalization, making it especially suited for diverse datasets where fixed thresholds may fall short. Our theoretical analysis and empirical results demonstrate AS-PCA's superiority in balancing variance explanation and sparsity, offering a powerful tool for modern high-dimensional data analysis.

4 - Robust Wafer Defect Pattern Classification and New Defect Pattern Detection in Imbalanced Data Using Few-Shot Learning

Byunghoon Kim, Hanyang University, Ansan, Korea, Republic of

In the semiconductor industry, Wafer Bin Maps (WBMs) generated through Electrical Die Sorting (EDS) inspection are crucial for identifying defect patterns on a wafer. The WBMs facilitate the classification of defect patterns based on the specific locations and distributions of faulty chips. Defect patterns often stem from specific manufacturing processes. Using machine learning models to quickly classify Wafer Bin Maps (WBMs) during production can greatly enhance process yield and product quality. Additionally, detecting new defect patterns is also crucial to improve model reliability.

In this study, we utilize the Prototypical Network, a few-shot learning technique, to classify existing defective patterns and detect new ones. This method effectively handles imbalanced data and requires minimal data for each class, making it particularly well-suited for scenarios with limited data availability. To reflect real manufacturing conditions, we adapted the existing model and incorporated class prototypes into the validation process without relying on labeled data. We developed a method for generating prototypes that excludes improperly labeled data from the dataset. By establishing class prototypes and calculating the distances between validation data and these prototypes, we effectively classified existing defect patterns and detected new ones. Compared to previous research, our approach demonstrated superior classification accuracy, particularly in significantly enhancing the accuracy for minority classes.

5 - Machine learning based business competitor identification

Nathaniel Smith, West Virginia University, Morgantown, WV, United States, Jeongsub Choi

Competitor identification (CI) is an essential step for successful business strategy development by strategically navigating the business environment and making informed decisions. Due to the complexity, various approaches to identifying competitors have been studied in the literature to shed light on the blind spots of managerial radars. Recently, modeling with machine learning has been actively adopted for efficient CI with vast amounts of data. In this talk, we present experimental results from predictive models for CI based on business organizational profiles.

SC64

Regency - 602

Social media, Transfer Learning, and AI

Invited Session

Social Media Analytics

Chair: Wendao Xue, The University of Texas at Austin, Austin, United States

1 - What to Sell and When to Sell in Live Streaming? An Online Decision-Making Approach

Jingwen Zhang, University of Washington, Seattle, WA, United States, Shaohui Wu, Yong Tan

Livestream selling has emerged as a popular and influential e-commerce format, yet little is known about how hosts effectively sell products in this dynamic environment. This study investigates the decision-making processes of hosts in livestream selling, focusing on their product assortment and presentation timing strategies. We propose a novel approach that combines the online learning framework of Thompson Sampling and the COX proportional hazards model to capture hosts' learning and decision-making behaviors. Specifically, we employ Thompson Sampling to model how hosts learn consumer preferences, predict demand, and decide on the product assortment to present during their livestreams. Concurrently, we utilize the COX proportional hazards model to examine the factors influencing hosts' decisions on the timing of each product set presentation. By applying these models to a dataset from one of the largest livestreaming platforms in China, we demonstrate how our approach effectively captures hosts' product assortment and presentation timing decisions in real practice. Our findings provide valuable insights into the learning and decision-making processes of hosts in livestream selling and offer managerial implications for enhancing the effectiveness of this increasingly important e-commerce format.

2 - Using Noisy Interval Data in Information System Research: Theory and Applications

Wendao Xue, University of Texas at Austin, Austin, TX, United States, Huidi Ma, Yifan Yu

This paper examines inference in regression models where one latent variable is known to belong to an interval with a certain probability, while other variables are measured accurately. We demonstrate that, under certain assumptions, the parameters of interest constitute an identified set. We propose an estimator for this identified set and show that it can be consistently estimated. Two applications are presented: (1) regression using variables generated by large language models, and (2) protecting information privacy when sharing data. For each application, we outline specific tailored assumptions, discuss the identification results, and detail the estimation process. The efficacy of our proposed method is demonstrated through simulations in each application. Additionally, we employ semi-simulated data to further validate the effectiveness of our method in these applications. This research illustrates that utilizing noisy interval data can enhance current methodologies, facilitating more accurate results under less stringent assumptions and reducing potential biases.

3 - Emotion AI in Disguise Spurs Strategic Behavior in Customer Care

Yu Kan, University of Washington- Michael G. Foster School of Business, Seattle, WA, United States

As organizations increasingly turn to AI for addressing customer complaints, understanding the dynamics of user interactions with AI and human agents becomes imperative. In this study, we investigate customers' strategic behavior when interacting with agents of different identity cues on a food delivery platform. We employ two AI-human interactive experiments, where participants are randomly assigned to AI agents (including a general AI and a GPT-based AI agent), a human agent, or a non-disclosure agent group. We observe significant strategic behavior and decreased satisfaction only in the non-disclosure group. The participants in the non-disclosure group present a unique paradox. Even though they report experiencing less negative emotions, their expressions are more negative, and they have the lowest satisfaction ratings among all groups. Our findings contribute to the growing research on AI-human interactions and emotion regulation, suggesting intriguing dynamics when the agent's identity is non-disclosed. This work provides valuable insights for organizations considering AI adoption for customer service, highlighting the potential challenges and implications for user experience, customer satisfaction, and solution acceptance.

SC66

Regency - 604

Frontiers in Causal Inference

Invited Session

Artificial Intelligence

Chair: Fei Fang, Yale University, New Haven, CT, United States

Co-Chair: Ruoxuan Xiong, Emory University, Atlanta, GA, United States

1 - Long-Term Causal Inference Under Persistent Confounding via Data Combination

Yuhao Wang, Tsinghua University, Beijing, China, People's Republic of, guido Imbens, Nathan Kallus, Xiaojie Mao

<blockquote mathjax"="">We study the identification and estimation of long-term treatment effects when both experimental and observational data are available. Since the long-term outcome is observed only after a long delay, it is not measured in the experimental data, but only recorded in the observational data. However, both types of data include observations of some short-term outcomes. In this paper, we uniquely tackle the challenge of persistent unmeasured confounders, i.e., some unmeasured confounders that can simultaneously affect the treatment, short-term outcomes and the long-term outcome, noting that they invalidate identification strategies in previous literature. To address this challenge, we exploit the sequential structure of multiple short-term outcomes, and develop three novel identification strategies for the average long-term treatment effect. We further propose three corresponding estimators and prove their asymptotic consistency and asymptotic normality. We finally apply our methods to estimate the effect of a job training program on long-term employment using semi-synthetic data. We numerically show that our proposals outperform existing methods that fail to handle persistent confounders.

2 - Regression Analysis for Conditional Spillover Effects Under Design-Based Uncertainty

Fei Fang, Yale University, New Haven, CT, United States, Laura Forastiere , Edoardo Airoldi

When interference exists, estimating spillover effects to correct bias for direct effects or the effects themselves is of interest. While regression analysis is commonly utilized for this task, e.g., regressing the outcome on the neighboring treatments, the analysis often relies on strong assumptions about the outcome model, such as homogeneity and additivity of the spillover effects. Alternatively, the weighted least squares (WLS) estimator from regression, with appropriately designed weights and covariate matrix, can be equivalent to the Hajek estimator for average spillover effects, which is model-free.

In this study, we focus on regression analysis for conditional spillover effects which are useful for understanding the heterogeneity of such effects across different groups. We propose integrated weights for the WLS estimator, demonstrating its equivalence to the Hajek estimator when the conditional covariate is categorical, without imposing assumptions on potential outcomes. For conditioning on continuous covariates, we assume linear outcome models but with heterogeneous coefficients. Under this assumption and design-based randomness, we define interpretable estimands and obtain consistent WLS estimators. We establish central limit theorem and derive a cluster-robust variance estimator under partial interference assumption. If homogeneous spillover effects can be further assumed, then the spillover effects can be decomposed and obtained from the dyadic regression between one's outcome and others' treatments separately. We examine the relationship between these estimators and the Hajek estimator and provide a variance estimator accommodating repeated dyads. Finally, we apply these methods to Honduras datasets to evaluate the spillover effects of educational programs on villagers' adoption behaviors.

3 - When the Bartik instrument meets topic models: a principled approach to text-based causal inferences

Tong Guo, Duke University, Durham, NC, United States

We propose the Bartik instrument as a principled source of IV for text-based causal inferences. The proposed procedure uses the estimated topic distribution from topic models to construct the Bartik instrument, and provides a useful alternative when the traditional BLP or Waldfogel IV does not work. We illustrate our approach in a study of the impact of social media news on Impossible Meats adoption by local restaurants and stores. Understanding this connection is challenging because of the lack of empirical measurement of local entrepreneur decisions at scale and, more importantly, the endogeneity of marketing communication to unobserved local demand shocks. We devise a unique location-specific adoption metric based on social media announcements, tracking local business decisions from 2015 to 2019. This metric is linked to comprehensive marketing communication extracted from social media using Natural Language Processing. We leverage the quasi-random variations of county-quarter-level news production for different topics to causally identify the linkage between social media publicity and adoption. We find that local coverage of social media news on the innovation increases the adoption of impossible meat products by local entrepreneurs. Interestingly, news content about producer financials was as important as content about sustainability and taste in driving local adoption of impossible meat products, potentially due to the signaling role of financial news for the trustworthiness of the technology (thus lowering uncertainty) and the trendiness of the technology (thus providing free marketing to small businesses who adopted the innovation).

4 - Data Integration for Efficient Causal Inference

Harsh Parikh, Johns Hopkins University, Durham, NC, United States, Elizabeth Stuart, Kara Rudolph

Recent works in the literature have focused on data integration from various studies for efficient causal effect estimation. These works typically assume that the intervention and outcome measures are same across all studies. In our work, we explore scenarios where the outcome measures are disparate across studies. We theoretically and empirically investigate when and how integrating data with disparate outcome measures can yield performance enhancement. Our findings indicate that performance enhancement is possible either under strong assumptions relating various outcome measures or on the existence of a calibration data where various outcome measures are simultaneously observed. We apply our data integration approach to study the effect of medication for treating opioid use disorder on the strength of withdrawal symptoms.

SC67

Regency - 605

Daniel H. Wagner Competition II

Award Session

Daniel H Wagner

Chair: James Cochran, The University of Alabama, Tuscaloosa, AL, United States

1 - Redesigning Zoning Systems for Equitable and Efficient Last-Mile Delivery at Ninja Van

Stanley Lim, Michigan State University, East Lansing, MI, United States, John Carlsson, Sheng Liu, Han Yu, Witsanu Arntong, Ee Hsin Tan

Last-mile logistics poses many challenges and high costs for firms. We develop a zoning system for Ninja Van to enhance last-mile delivery efficiency and boost customer satisfaction. By incorporating Voronoi Diagrams, stochastic and robust optimization techniques, we introduce a novel zoning optimization procedure and tested the model's efficacy in a major Southeast Asian city. The results showed substantial reduction in delivery work spans and the time drivers spend on deliveries. We project yearly savings exceeding $400,000 just by redefining zones and their boundaries in one country.

2 - Transportation marketplace rate forecast using signature transform

Haotian Gu, UC Berkeley, Albany, CA, United States, Xin Guo, Timothy Jacobs, Philip Kaminsky, Xinyu Li

This work develops a novel statistical method leveraging signature transforms to predict freight transportation marketplace rates. Our approach utilizes the universal nonlinearity property of signature transform to linearize the feature space and hence translates the forecasting problem into linear regression, and uses signature kernels for efficient comparison of time series data. This enables precise feature generation and identification of regime switching. Deployed by Amazon trucking operations, our algorithm surpasses industry models by improving prediction accuracy by over fivefold.

SC68

Regency - 606

Undergradutate Operations Research Prize II

Award Session

Undergraduate Operations Research Prize

Chair: Rachel Cummings, Columbia University, New York, NY, United States

1 - Planning Adaptive Experiments: A Mathematical Programming Approach

Jimmy Wang, Columbia University, New York, NY, United States, Ethan Che

Standard adaptive experimentation algorithms overlook important practical issues including multiple objectives, non-stationarity, batched feedback, constraints, and personalization. Moving away from developing bespoke algorithms for each setting, we present a mathematical programming framework that flexibly incorporates a wide range of objectives, constraints, and statistical procedures. By formulating a dynamic program in the batched limit, our modeling framework enables the use of scalable optimization methods (e.g., SGD and auto-differentiation) to solve for treatment allocations.

2 - The maximum length car sequencing problem

Lara Pontes, Universidade Federal da Paraíba, João Pessoa, Brazil, Carlos Neves

The maximum length car sequencing problem aims to support the assembly operations arising in multinational automotive companies. We propose an integer programming formulation to schedule the maximum number of cars without violating spacing constraints associated with car options (sunroof, radio), a valid combinatorial upper bound, and iterative algorithms to solve the problem when good primal bounds are not available. An ILS-based heuristic enhances the performance of the exact methods. Computational results achieve low gaps for benchmark instances and optimally solve the company's demands.

3 - A Projection-Free Method for Solving Convex Bilevel Optimization Problems

Khanh-Hung Giang-Tran, University of Sydney, Sydney, Australia

We consider a class of convex bilevel optimization problems, where we minimize a convex smooth objective function over the optimal solution set of another convex smooth constrained optimization problem. We propose a new projection-free method for convex bilevel optimization requiring only a linear optimization oracle over the domain. We establish $O(t^{−1/2})$ convergence rate guarantees in terms of both inner- and outer-level objectives and demonstrate how additional assumptions result in accelerated rates of up to $O(t^{−1})$ and $O(t^{−2/3})$ for inner- and outer-levels respectively.

4 - Allocation of Surveillance Assets in Undersea Warfare

Sebastian Martin, United States Naval Academy, Annapolis, MD, United States

Recent advances in submarine technology require an effective national strategy for detecting undersea threats. We formulate an optimization model to effectively deploy sensors, maximizing the expected number of detected targets based on historical threat data and sensor resources. The model accommodates sensor overlap, introducing nonlinearity, which we handle by linearizing with a logarithmic transformation and tangent line approximations. We solve the model with notional sensor and maritime shipping traffic data.

SC69

Regency - 607

Impacts of Emerging Technologies

Invited Session

eBusiness

Chair: Jiali Zhou, American University, Washington DC, DC, 200016, United States

1 - Navigating Product Diversification in Live Streaming E-Commerce: Evidence from Douyin

Jingyun Hu, Clemson University, Pendleton, SC, United States, Yi Gao, Keran Zhao

The past decade has witnessed the growing prevalence of live streaming selling (LSS). As streamers serve as a proxy between manufacturers and consumers, product assortment management has been a critical strategic factor for their success. Relying on data collected from Douyin, a prominent Chinese LSS platform, this study aims to investigate the impact of diversification on LSS performance through the lens of adding new categories in live streaming sessions. The results in this study show that product category diversification, in general, has a significant positive effect on sales during live streaming events. However, this effect is alleviated by the semantic similarity between the newly introduced and existing categories. In addition, we find that product diversification has a negative impact on the sales of incumbent categories. This study contributes to the literature on product assortment by providing insights into the role of product diversification in the context of real-time selling. We also discuss the managerial implications for the platform and streamers.

2 - How Does Popularity Information Affect Product Design?

Guangrui Li, York University, Toronto, ON, Canada, Zheng Gong, Zhepeng Li

Popularity information serves as quality signal for consumers to learn about products. Past literature has shown that the revealing popularity information may herd consumers to the popular products and leads to a superstar phenomenon, and narrow-appeal products benefit more from the same popularity. Up to our knowledge, the previous literature takes the product design as given, but the firms could change their product design in response to the popularity information revealing. In this paper, we examine the impact of popularity information revealing on firms' product design strategy by using a policy change on Wechat Official Account platform that reveals articles' popularity to subscribers. More specifically, we are interested in understanding the following aspects: (1) The quality and amount of advertisement in the content; (2) The topic choice of the contents. Our results show that the official accounts tend to increase their topic diversity as a strategic response to the demand shock induced by the reveal of popularity information.

3 - How Generative AI Shapes Video Sharing Platform: a Tale of Two Forces

Luying (Iris) Qiu, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Oliver Wei, Jiali Zhou, Weiyin Hong, Kai-lung Hui

We study the impact of generative AI on participants of video-sharing platforms using a natural experiment setting: the introduction of AI summary feature on one of the largest video-sharing platforms in China in November 2023. By leveraging a regression discontinuity strategy, we find that AI summaries significantly reduced creators’ contribution, viewers’ in-video engagement, and the number of reviews. We also find that the reduction in creators’ contribution is more severe for inexperienced creators. We draw related research and managerial implications.

4 - Logistics Service Sharing in Cross-Border E-Commerce

Nevin Mutlu, Eindhoven University of Technology, Eindhoven, Netherlands, Faranak Khooban, Ton de Kok

As demand for cross-border e-commerce has grown rapidly, challenges have emerged for both retailers and consumers participating in this global market. Retailers have been struggling with high logistics costs to fulfill cross-border demand, while they also suffer from consumers' lack of trust in foreign retailers. In this paper, we study a cross-border collaboration scheme between a domestic and a foreign retailer to mitigate these challenges. This entails a co-opetition framework where the domestic retailer is responsible for the last-mile delivery of the foreign retailer's orders in exchange for a logistics service fee. We model demand via an MNL choice model incorporating trust- and price-sensitive consumers. We compare the market outcomes of the two retailers in "pre-collaboration" and "post-collaboration" settings. We find that there exist win-win outcomes where both retailers benefit from collaboration under realistic settings. We also show that a cooperative mechanism can lead to higher profits for both retailers compared to the non-cooperative mechanism for setting the logistics service fee, if the contract terms are decided carefully.

SC70

Regency - 701

Optimization Approaches to Enhancing Biodiversity

Invited Session

Energy: Natural Resources

Chair: Sabah Bushaj, SUNY, Plattsburgh, NY, 12901, United States

1 - Optimal Sampling Strategy for Probability Estimation: An Application to the Agricultural Quarantine Inspection Monitoring Program

Huidi Ma, The University of Texas at Austin, Austin, TX, United States, Benjamin Leibowicz, John Hasenbein

Imported agricultural pests can cause substantial damage to agriculture, food security, and ecosystems. In the United States, the Agricultural Quarantine Inspection Monitoring (AQIM) program conducts random sampling to estimate the probabilities that cargo and passengers arriving at ports of entry carry pests. Assessing these risks accurately is critical to enable ef- fective policies and operational procedures. In this paper, we formulate an optimization model that minimizes the mean squared error of the probability estimates that AQIM obtains. The central decision-making tradeoff that the model explores is whether it is preferable to sample more arriving containers (and fewer boxes per container) or more boxes per container (and fewer containers), given limited resources. We first derive an analytical solution for the optimal sam- pling strategy by leveraging several approximations. Then, we apply our model to a numerical case study of maritime cargo sampling at the Port of Long Beach. We find that, across a wide range of parameter settings, the optimal strategy samples more containers (but fewer boxes per container) than the current AQIM protocol. The difference between the two strategies and the accuracy improvement with the optimal approach are larger if the pest statuses of boxes in the same container are more strongly correlated.

2 - Managing Large-Scale Invasions: Simulation-Optimization with Gaussian Dispersal Kernels and Stochastic Seed Establishment

Sevilay Onal, University of Illinois Springfield, Springfield, IL, United States

Biological invaders cause substantial economic losses exceeding $21 billion annually for the US government by damaging crops, agricultural lands, habitats, and sustainability. Aggressive invaders recognized under the Federal Noxious Weed Act of 2000, such as Sericea lespedeza, drive continuous federal efforts to prevent, eliminate, and manage invasive species. This highlights the critical need for optimal control methods to minimize their detrimental effects on biodiversity and the economy. To our knowledge, this paper is the first to integrate completely random occurrences of an invader that are not explainable by biophysical impacts within an integrated simulation- optimization model to control Sericea. The simulation model estimates the natural behavior of seed dispersal using Gaussian cell-to-cell transition probabilities and an algorithm developed to estimate the random or unexpected behavior of the invader. The case study data and parameter calibration are based on large-scale field data collected in Kansas and Oklahoma. We simulate Sericea growth over a 2500-acre landscape for 25 years and optimize search and treatment locations while minimizing its economic damage to forage production under a restricted budget. The results of the simulation-optimization framework provide information on the optimal treatment location and frequency, the optimal search speed, and the cost-benefit analysis of treatment in various invasion and seed dispersal scenarios. The computational results suggest that more budget is allocated to search in lower-density areas than treatment. If the weed is detected in a cell, the entire area is searched for a possible stochastic establishment. In higherdensity scenarios, treatment is preferred, and a search

3 -

An Integer Programming Approach for Designing Wildlife Conservation Corridors with Geometry Requirements

Chao Wang, Arizona State University, Tempe, AZ, United States, Jorge Sefair

Conservation corridors have been widely used to prevent biodiversity loss by linking fragmented habitats. One of the key design challenges involves identifying corridors that meet ecological requirements, including considerations of connectedness, width, and length. In this study, we propose a MIP model to enforce geometry requirements into the corridor design problem, aiming to maximize the corridor’s utility within a grid landscape. To enhance solution performance, we develop a Branch-and-Cut algorithm, paired with a bound-improving heuristic and a variable reduction preprocessing algorithm. A byproduct of our work is a polynomial-time algorithm for the solution of the minimization variant of this problem, focusing on finding a least-cost corridor with a specified width. Our algorithm is tested on a real instance for the protection of the Florida panther and a set of computer-generated landscapes. Test results demonstrate the algorithm’s ability to find near-optimal solutions for landscapes comprising up to 3000 patches within a two-hour timeframe.

4 - An Adaptive Simulation-Optimization Framework to Survey and Control Invasive Species

Sabah Bushaj, SUNY, Plattsburgh, NY, United States

Invasive species are a serious threat to ecosystems worldwide, calling for innovative methods to survey and control them effectively. This research presents a new approach—an adaptive simulation-optimization framework—specifically designed to tackle the complexities of managing invasive species. This framework is highly adaptable, providing standardized features for different species and offering various simulation options to match different environments and species dynamics. Researchers and practitioners can use this framework to tailor simulations for accurate surveys and strategic control efforts. By incorporating advanced solvers and allowing for problem-specific adjustments, it ensures reliable optimization customized to the specific challenges of invasive species scenarios.

SC71

Regency - 702

Advanced Algorithms for Machine Learning and Discrete Optimization

Invited Session

Data Mining

Chair: Yongchun Li, Georgia Institute of Technology, Atlanta, GA, United States

1 - Convergence of Gradient Descent with Small Initialization for Unregularized Matrix Completion

Salar Fattahi, University of Michigan, Ann Arbor, MI, United States, Jianhao Ma

We study the problem of symmetric matrix completion, where the goal is to reconstruct a positive semidefinite matrix of low rank from only a subset of its observed entries. For the first time, we prove that vanilla gradient descent (GD) with small initialization converges to the ground truth, without requiring any explicit regularization or projection. This convergence result holds true even in the over-parameterized scenario, where the true rank is unknown and conservatively over-estimated. It also achieves a near-optimal sample complexity. The existing results for this problem either require explicit regularization, a sufficiently accurate initial point, or exact knowledge of the true rank. At the crux of our method lies a novel weakly-coupled leave-one-out analysis, which allows us to establish the global convergence of GD, extending beyond what was previously possible using the classical leave-one-out analysis.

2 - Derivation and Generation of Path-Based Valid Inequalities for Transmission Expansion Planning with New Bus Integration

Behnam Jabbari Marand, North Carolina State University, Raleigh, NC, United States, Adolfo Escobedo

This research tackles an extension of the DC OPF-based transmission expansion planning problem where both new lines and buses can be added to the existing network. To handle the general intractability of this problem, the primary computational strategy involves deriving valid inequalities (VIs) that leverage the problem's structure. Effective VIs are generated by considering different paths within the existing and expanded network both a priori and during the solution of the dispatch problem. While in the associated theorems identifying the longest paths is necessary to ensure validity in the general case, the proposed approach circumvents such an inefficient procedure for the majority of bus pairs. It is shown that the path-based VIs derived through this method dominate those obtained from solving the longest path problem, leading to a stronger problem formulation. This work also proposes a methodological framework for constructing an effective cut pool and identifying strong VIs to generate cutting planes. Two classes of VIs result from these insights, and their efficacy and computational advantages are demonstrated through application to a modified Polish 2383-bus system.

3 - Statistically Optimal K-Means Clustering via Nonnegative Low-Rank Semidefinite Programming

Richard Zhang, University of Illinois, Urbana, IL, United States, Yubo Zhuang, Xiaohui Chen, Yun Yang

K-means clustering is a widely used machine learning method for identifying patterns in large datasets. Semidefinite programming (SDP) relaxations have recently been proposed for solving the K-means optimization problem that enjoy strong statistical optimality guarantees, but the prohibitive cost of implementing an SDP solver renders these guarantees inaccessible to practical datasets. By contrast, nonnegative matrix factorization (NMF) is a simple clustering algorithm that is widely used by machine learning practitioners, but without a solid statistical underpinning nor rigorous guarantees. In this paper, we describe an NMF-like algorithm that works by solving a nonnegative low-rank restriction of the SDP relaxed K-means formulation using a nonconvex Burer--Monteiro factorization approach. The resulting algorithm is just as simple and scalable as state-of-the-art NMF algorithms, while also enjoying the same strong statistical optimality guarantees as the SDP. In our experiments, we observe that our algorithm achieves substantially smaller mis-clustering errors compared to the existing state-of-the-art.

4 - Integer Programming for Learning Directed Acyclic Graphs from Non-Identifiable Gaussian Models

Tong Xu, Northwestern University, Evanston, IL, United States, Armeen Taeb, Simge Kucukyavuz, Ali Shojaie

We study the problem of learning directed acyclic graphs from continuous observational data, generated according to a linear Gaussian structural equation model. State-of-the-art structure learning methods for this setting have at least one of the following shortcomings: i) they cannot provide optimality guarantees and can suffer from learning sub-optimal models; ii) they rely on the stringent assumption that the noise is homoscedastic, and hence the underlying model is fully identifiable. We overcome these shortcomings and develop a computationally efficient mixed-integer programming framework for learning medium-sized problems that accounts for arbitrary heteroscedastic noise. We present an early stopping criterion under which we can terminate the branch-and-bound procedure to achieve an asymptotically optimal solution and establish the consistency of this approximate solution. In addition, we show via numerical experiments that our method outperforms three state-of-the-art algorithms and is robust to noise heteroscedasticity, whereas the performance of the competing methods deteriorates under strong violations of the identifiability assumption. The software implementation of our method is available as the Python package micodag.

SC72

Regency - 703

Socially and Environmentally Responsible Operations

Invited Session

MSOM: Sustainable Operations

Chair: Vibhuti Dhingra, Schulich School of Business, York University, Toronto, ON, Canada

Co-Chair: Ali Kaan Tuna, Tilburg University, Tilburg, N/A, Netherlands

1 - Managing Carbon-Neutral Delivery in Online Retailing

Huseyn Abdulla, University of Tennessee, Knoxville, TN, United States, Seulchan Lee, Han Oh

Order-delivery related emissions constitute the vast majority of the environmental footprint of online retailers. A growing segment of consumers has become concerned with these emissions, discouraging them from shopping online. To alleviate these concerns, many online retailers consider committing to carbon-neutrality in the order delivery domain. These retailers partner with Sustainability as a Service (SaaS) providers to offer a voluntary, carbon offset-based green shipping option to eco-conscious consumers to reduce the cost burden of offsetting the emissions. Using a game theoretical approach, we examine the conditions under which it is optimal for a profit-maximizing online retailer to become carbon-neutral in the order delivery domain and to offer a consumer-paid green shipping option by partnering with a SaaS provider. We compare two common offset payment models that back the green shipping option, namely, preset and calculated offset payment models, and explore the implications of a hybrid approach that combine the advantages of both models.

2 - An Operational Perspective on the Role of Liability in Microfinancing

Elaheh Rashidinejad, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Opher Baron, Gonzalo Romero

We study the effects of different liability structures on the microfinancing performance in low- and middle- income countries through an operational lens. We investigate this problem using a Newsvendor with financing and effort framework. The Base of the Pyramid entrepreneur's outcome depends on her production quantity, effort level and business characteristics. The entrepreneur's effort directly impacts the success of her business in a multiplicative way. Furthermore, she starts with zero initial budget and borrows a loan to operate her business.One extreme of the liability structure we consider is a social bank that operates as an external lending institution and cannot collect the entire debt in the presence of business losses. The other extreme is a community bank that consists of peers in a community gathering their savings to form a bank and using their social ties to collect the debt eventually. We study these banks under profit-maximizing or zero-profit objectives, where the banks may face bankruptcy costs. Our study reveals that under any of these microfinancing setups, the effort exerted by the entrepreneur is the main driver of her equilibrium utility. Further, the effectiveness of microfinancing depends on two main factors: the bankruptcy cost and the product's critical fractile. Surprisingly, we show that the peer pressure mechanism may not always benefit both the entrepreneur and the bank. We establish conditions under which the zero-profit (profit-maximizing) community bank generates higher social welfare than the zero-profit (profit-maximizing) low-liability banks. A numerical study provides additional insights on the performance of moderate-liability banks.

3 - Responsible Operations in the Mining Sector: Do Penalties Improve Safety?

Vibhuti Dhingra, Schulich School of Business, York University, Toronto, ON, Canada, Anna Saez de Tejada Cuenca

Mining is one of the most hazardous industries in the world. Workplace injury rates in mining are 14.2 for every 100,000 workers—273% higher than other sectors, and the industry had more than 500 fatalities between 2017-2022. With the expected growth in mineral extraction necessary to power a green economy, the sector is only poised to expand further.

Given their massive footprint, mines are regularly inspected for compliance with workplace safety and health standards. The Mine Safety and Health Administration (MSHA) in the U.S., for example, conducts over 12,000 inspections every year. If a violation is detected, the inspectors enforce corrective measures and levy financial penalties on the responsible entities. In 2010, for instance, the MSHA imposed 160 million dollars in fines for mine safety violations.

Penalties can act as a deterrent against future violations by prompting mine operators to make process improvements such as enhancing ventilation and ensuring proper machine maintenance. However, penalties can also be ineffective if mine operators do not pay the finesand instead continue business as usual.

What is the causal effect of penalties on mine safety? In this paper, we answer this research question using data on mine safety inspections, violations, and penalties from the MSHA—spanning over one million inspections conducted on more than 4,000 coal and 18,000 metal/non-metal mines in the U.S. involving over 14,000 mine operators. Our identification strategy leverages the passage of the MINER Act in 2006 that significantly increased the penalties imposed for safety violations.

4 - Outsourcing, Reshoring, and Carbon Emissions in Supply Chains

Ali Kaan Tuna, Tilburg University, Tilburg, Netherlands, Robert Swinney

Nearshoring has garnered significant attention from firms, governments, and policymakers in recent years, and has emerged as one of the most prominent supply chain strategies to build more sustainable (and resilient) supply chains. In decentralized supply chains, an agent's reshoring decision might have an important impact on the whole supply chain. Motivated by the potential economic and environmental consequences of nearshoring in decentralized supply chains, we study a two-tier decentralized supply chain, where a buyer needs to source a critical component from a supplier to manufacture a single end-product.

Well before the buyer makes any procurement or production decisions, the firms first decide on their locations to form a supply chain to manufacture, with the buyer acting as a leader in this phase. They can form a fully offshore supply chain, a hybrid supply chain with the buyer onshore and the supplier offshore, or a fully nearshore supply chain. Firms' reshoring decisions change their fixed and variable costs, the per-unit environmental impact of their production and distribution processes, and when the buyer has to choose a procurement quantity for the component or a production quantity for the end-product. Our chief goal is to understand when the firms form each type of supply chain configuration in equilibrium, and the environmental implications of the equilibrium outcome. We also seek to understand how the relative power of the firms in the supply chain has an impact on their reshoring decisions and the subsequent environmental impact.

SC73

Regency - 704

MSOM Student Paper Competition Finalists

Award Session

Manufacturing and Service Operations Management (MSOM)

Co-Chair: Vasiliki Kostami, HEC Paris, Jouy-En-Josas, France

1 - Optimizing Health Supply Chains with Decision-Aware Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States

We address the problem of allocating limited medical resources in a developing country by combining ML (to predict demand) with optimization (to optimize allocations). A key challenge is aligning the ML model's loss function with the decision loss in optimization. We propose a scalable decision-aware learning framework and successfully deployed in collaboration with the Sierra Leone government across 1,123 healthcare facilities nationwide, leading to a 15-29% increase in medicine consumption and improved real-world patient access to care.

2 - Signaling Competition in Two-Sided Markets

Yuri Fonseca, Stanford University, Stanford, CA, United States

We consider decentralized platforms facilitating many-to-many matches between two sides of a marketplace. In the absence of direct matching, inefficiency in market outcomes can easily arise. For instance, popular supply agents may garner many units from the demand side, while other supply units may not receive any match. A central question for the platform is how to manage congestion and improve market outcomes. We study the impact of a detail-free lever: the disclosure of information to agents on current competition levels. Disclosing competition reduces the perceived value of popular units, but, at the same time, it can help agents on the other side better elect across options. How large are such effects, and how do they affect overall market outcomes? We answer this question empirically. We partner with the largest service marketplace in Latin America, which sells non-exclusive labor market leads to workers. We propose a structural model which allows workers to internalize competition at the lead level and captures the equilibrium effect of such reaction to competition at the platform level. We estimate the model by leveraging agents' exogenous arrival times and a change in the platform's pricing policy. Using the estimated model, we conduct counterfactual analyses to study the impact of signaling competition on workers' lead purchasing decisions, the platform's revenue, and the expected number of matches. We find that signaling competition is a powerful lever for the platform to reduce congestion, redirecting demand, and ultimately improving the expected number of matches for the markets we analyze.

3 - Dynamic Matching with Post-allocation Service and its Application to Refugee Resettlement

Soonbong Lee, Yale university, New Haven, CT, United States

Motivated by our collaboration with a major refugee resettlement agency in the U.S., we study a dynamic matching problem where each new arrival (a refugee case) must be matched immediately and irrevocably to one of the static resources (a location with a fixed annual quota). In addition to consuming the static resource, each case requires post-allocation services from a server, such as a translator. Given the uncertainty in service time, a server may not be available at a given time, thus referred to as a dynamic resource. Upon matching, the case waits to avail service in a first-come-first-serve manner. Bursty matching may result in undesirable congestion of the servers. Consequently, the planner (the agency) faces a dynamic matching problem with an objective combining the matching reward (pair-specific employment outcomes) with the cost for congestion for dynamic resources and over-allocation for the static ones. Motivated by the observed fluctuations in the refugee pools across the years, we aim to design algorithms with no distributional knowledge. We develop learning-based algorithms that are asymptotically optimal in certain regimes, easy to interpret, and computationally fast. Our design is based on learning the dual variables of the underlying optimization problem; however, the main challenge is the time-varying nature of the dual variables associated with dynamic resources. Our theoretical development integrates techniques from Lyapunov analysis, adversarial online learning, and stochastic optimization. When tested on our partner agency's data, our method outperforms existing ones making it a viable candidate for replacing the current practice upon experimentation.

4 - Causal Message Passing: A Method for Experiments with Unknown and General Network Interference

Sadegh Shirani, Stanford University, Stanford, CA, United States

Randomized experiments are a powerful methodology for data-driven evaluation of decisions. Yet, their validity may be undermined by network interference. This study introduces a new framework to accommodate complex and unknown network interference. Our framework, termed causal message-passing, is grounded in approximate message passing methodology. Utilizing causal message-passing, we introduce a practical algorithm to estimate the total treatment effect. We demonstrate the effectiveness of this approach across five numerical scenarios, each characterized by a distinct interference structure.

SC74

Regency - 705

ENRE: Risk Analysis for Electricity Markets

Invited Session

ENRE: Electricity

Chair: Ryan Ent, University of Massachusetts-Amherst, Ringoes, NJ, 08551, United States

Co-Chair: Golbon Zakeri, University of Massachusetts - Amherst, Amherst, MA, United States

1 - Analysis of ISO New England's Energy Imbalance Reserve Product with Risk Averse Agents

Ryan Ent, University of Massachusetts-Amherst, Amherst, MA, United States, Golbon Zakeri

In wholesale electricity markets, operating reserves are contracts for generation capacity sold by electricity providers to provide the power system with flexibility in the case of N-1-1 contingencies. In 2025, the independent system operator of New England(ISO-NE) will introduce a co-optimized day ahead energy and reserve market that contains four reserve products with a call option settlement based on the difference between the real time price of energy and a strike price set by ISO-NE. Part of this market is a new product, the energy imbalance reserve, which will make up the difference between energy cleared on the day ahead market and the system operator's real time load forecast. ISO-NE's goals for this market are to incentivize the generators in their fleet, mainly reliant on natural gas, to invest in advanced fuel and increase energy purchased on the day ahead to a level equal to the expected real time load. To analyze how this market may or may not achieve these goals, we model the simultaneous equilibrium of multiple risk-averse generators participating in day ahead and real time markets for energy with and without the presence of energy imbalance reserve.

2 - ISO New England Regional Energy Shortfall Threshold Project

Jinye Zhao, ISO New England, Holyoke, MA, United States

ISO New England is pioneering the development of a Regional Energy Shortfall Threshold (REST), establishing an acceptable level of regional energy shortfall risk during extreme weather conditions. The development of REST will utilize the Probabilistic Energy Adequacy Tool, which is capable of identifying weather events that could cause energy deficits as well as assessing the likelihood and severity of such events. REST will serve as an energy adequacy standard for extreme weather events, complementing the widely used 1-day-in-10-year resource adequacy criterion. This presentation focuses on ISO's current thinking regarding the selection criteria for extreme weather events, the metrics used to quantify energy adequacy risk and the frequency of REST evaluations.

3 - Towards Completion of Risk Markets

Iman Khajepourtadvani, University of Massachusetts Amherst, Amherst, MA, United States, Golbon Zakeri

The seminal paper of Ralph and Smeers (2011) demonstrates that when risk markets are complete, the competitive risk-averse equilibrium of a game, where agents are endowed with coherent risk measures is equivalent to a system optimization problem. This finding is of interest to policymakers in many markets, including energy markets. In this presentation, we will lay out some foundational steps for completing risk markets.

4 - The Impact of Capacity Accreditation on Long-term Resource Adequacy

KE XIN ZUO, Cornell University , Ithaca, NY, United States, Jacob Mays

The rapidly evolving global energy landscape reveals critical flaws in current resource adequacy (RA) mechanisms. Weak non-performance penalties in capacity markets incentivize suppliers to over-promise and under-deliver, leaving the grid vulnerable during extreme events. Central to this issue is capacity accreditation - the process by which ISOs assess each resource's contribution to system reliability during stress periods. A popular approach to assigning capacity credits is measuring generators' effective load carrying capability (ELCC) over designated performance assessment hours (PAHs). However, the misalignment of these PAHs with actual shortage events, along with inconsistencies in measuring shortfall risk, often lead to systemic over-accreditation. This paper first addresses the need for a standardized characterization of resource adequacy in wholesale electricity markets. Building on this foundation, we introduce an economically aligned ELCC and corresponding capacity accreditation framework that incentivizes optimal levels of capacity investment while meeting specified reliability standards. Using a stochastic optimization framework, we then compute the market equilibria arising under various approaches to accreditation for a numerical example, demonstrating the sub-optimality of standard practices. This study contributes to understanding how capacity accreditation affects long-term resource adequacy and informs the design of efficient market mechanisms and policies to ensure a resilient grid.

SC75

Regency - 706

Modeling the Impact of Climate Change on the Electric Power Sector: Inputs, Assumptions, and Use Cases

Invited Session

ENRE: Energy-Climate

Chair: Ana Dyreson, Michigan Technological University, Houghton

1 - Enhancing Climate Resilience in Power System Planning: An Integrated Framework

Maren Ihlemann, EPRI International, Golden, BC, Canada

Climate change has been manifesting in more frequent and extreme weather conditions which expose power systems to escalating challenges. These challenges are projected to intensify over the next decades. Concurrently, there's a growing dependency of electricity as main energy carrier for multiple sectors, increasing the need for a robust infrastructure.

Traditionally, power system planning was based on extrapolating historically observed conditions, with limited representation of weather data and visibility of climate-related hazards. However, modeling advances allow for more complex representations of hazards and their impacts in power system planning. The robustness of the design will depend on the resiliency considerations and metrics adopted, as well as the tolerance enforced by the planner to withstand stressful periods.

A holistic planning to operations solution requires an integrated planning framework that leverages climate data and asset vulnerabilities against adverse conditions. By connecting various power systems planning functions, key climate hazards are identified, adapted to, and ultimately mitigated. As a first step to this framework, an initial buildout is screened over a large set of climate-projected weather data identifying potential periods at risk using the Risk Screening tool (*RiSc*). Key potential risks are then passed to the capacity expansions tool, extended to explicitly incorporate climate hazards and adaptations. To assess expected system reliability, resource adequacy is assessed over a large set of operating conditions considering system weather dependencies.

The framework is applied to a synthetic Texas case study and results demonstrate that resilient planning significantly enhances system reliability and effectively mitigates extreme weather impacts.

2 - Generative Super Resolution for Estimating Climate Change Impacts on Power Systems and Communities

Brandon Benton, National Renewable Energy Laboratory, Golden, CO, United States, Grant Buster, Jordan Cox, Andrew Glaws, Ryan King

As climate change progresses, high-resolution climate-change-impacted weather data is crucial for assessing the impacts on power systems and communities. We introduce methods using generative adversarial networks (GANs) to downscale global climate model (GCM) data, achieving enhanced spatiotemporal resolutions suitable for power system and community analysis. We have effectively downscaled key power system modeling inputs, including wind, solar, temperature, and humidity, to 4km hourly resolutions. This advancement facilitates more nuanced power system analysis and better adaptation planning. Similar methods combine satellite imagery and human settlement layers to generate 1-km hourly urban heat island estimates. These estimates are crucial for evaluating heat risk and resilience in urban areas, including the impact of climate change on energy cost and burden, especially during future extreme events. By leveraging GANs, our approach significantly improves climate data resolution at low computational cost, offering new insights into the impacts of climate change on power infrastructure and urban communities, thus aiding in the development of more informed and resilient adaptation strategies.

3 - Navigating Climate Uncertainty: A Collaborative Approach for Grid Planning in California

Liyang Wang, University of California, Berkeley, Berkeley, CA, United States, Owen Doherty, Kripa Jagannathan, Nancy Freitas, Julia Szinai, Andrew Jones

The escalating frequency and intensity of climate change-driven extreme weather events expose critical vulnerabilities in our energy infrastructure. Concurrently, the rapid shift towards electrification and renewable energy introduces deep uncertainties into long-term grid planning. These interconnected stressors are challenging to address using traditional planning and modeling approaches. Planners are seeking actionable and computationally tractable approaches that explore how climate change impacts grid design and that examine the tradeoffs among the competing priorities among stakeholders. We describe a co-production effort between electricity sector practitioners, climate scientists, and power system modelers to translate climate projections into grid model inputs.

Using California as a case study, we define extreme events and establish appropriate thresholds using cluster analysis and extreme value theory, translating climate projections into demand scenario inputs for a capacity expansion model. We then introduce the application of Decision-Making under Deep Uncertainty (DMDU) framework to capacity expansion planning. We discuss how DMDU allows for multi-objective evaluation of different grid designs across a wide range of climate scenarios in a computationally tractable manner, enables planners to explore different definitions of robustness and grid resilience, and incorporates different risk tolerance from different stakeholders. This case study provides insight into the actionability of climate science and DMDU, illustrates the challenges in applying DMDU and coproduction, and bridges the gap between climate science projections and the needs of grid planners.

4 - What is the Contribution of Transmission to U.S. Resource Adequacy?

Charalampos Avraam, National Renewable Energy Laboratory, Golden, CO, United States, Jessica Kuna, Brian Sergi

Resource adequacy encompasses the availability of resources to cover future electricity demand. In a fuel-dominated electricity mix, resource adequacy metrics aimed to ensure sufficient future nameplate capacity. Fuel-fired generators can be sited in proximity to demand, and generation capacity offers predictable electricity delivery under normal operating conditions. Traditional resource adequacy frameworks either focus on ensuring generation capacity availability, or assess whether a system is resource adequate for some given generation capacity. However, nameplate capacity availability of fuel-fired generators does not always translate into electricity delivery. Fuel delivery to natural gas-fired plants failed during the 2021 Texas Power Crisis, rendering existing capacity inoperable. Moreover, the rapid deployment of energy-limited resources, including renewables and batteries, challenges existing resource adequacy frameworks and couples regional electricity demand and renewables potential via transmission. While ongoing research illuminates the contribution of transmission to resource adequacy, long-term planning fails to quantify transmission contributions, which are eventually disregarded in real-life decision-making. Our contribution is twofold. First, we go beyond the existing generation capacity-centered approaches to design metrics for the assessment of transmission contributions to resource adequacy. Second, we leverage a large-scale capacity expansion model within a multi-objective optimization framework to quantify the merits of transmission expansion under alternative decarbonization and policy pathways. We enhance the Regional Energy Deployment System (ReEDS) and generate Pareto Frontiers across engineering and economic resource adequacy metrics. This work aims to inform policy-makers and regional stakeholders engaged in long-term transmission planning under heavy penetration of energy-limited resources in highly decarbonized futures.

5 - A Universal Synthetic Weather Data Generator for Use in Power System Resource Adequacy, Short-term Optimization, and Control

Duc-Huy Pham, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, United States, Jordan Kern

Both long-term capacity expansion and short-term operational models of power systems require some characterization of weather-based uncertainties in electricity supply and demand. Resource adequacy, defined as the ability of the electric grid to satisfy end-user power demand at any time, is typically measured by subjecting models to past observations (weather reanalysis), outputs from forward-looking global climate models (GCMs), or a combination of both. To ensure more robust outcomes, system operators should evaluate planning and operational strategies using very large samples. However, when using weather reanalysis or global climate datasets, increasing the sample size pulls data from a wider temporal range, potentially obscuring meaningful shifts in the underlying system state (e.g., due to climate change). Here, we present a fairly universal approach for creating synthetic meteorological ensembles (including short-term forecasts) of unlimited size that replicate the statistical and time series characteristics of any user-defined "base" dataset (e.g., a 20-year window of GCM outputs or a 10-year window of historical observations). We demonstrate how this approach can be used in multiple modeling applications (resource adequacy, short-term optimization, and control) and how it can be easily transferred to any U.S. balancing authority.

SC76

Regency - 707

Critical Mineral and Lithium Supply Chains for Clean Energy Technologies

Invited Session

ENRE: Other Energy

Chair: Erick Jones, University of Texas at Arlington, Arlington, TX, United States

1 - Lithium Supply Chain Optimization

Sarasadat Alavi, University of Texas at Arlington, Arlington, TX, United States

The surging demand for lithium in the United States necessitates the development of efficient and sustainable supply chain strategies. This research introduces a novel lithium closed-loop supply chain model that optimizes planning for different regions in the USA while considering uncertainty. The model employs stochastic programming to address inherent uncertainties such as demand fluctuations and resource availability. Capacity acquisition decisions are incorporated to ensure the supply chain can meet the growing demand for lithium. The model emphasizes sustainability, seeking to minimize the ecological footprint of the lithium supply chain. Social fairness is also addressed, ensuring the equitable distribution of benefits and burdens among different communities and stakeholders. This research contributes to the development of a more resilient, sustainable, and socially responsible lithium industry in the United States by integrating critical aspects into a comprehensive closed-loop supply chain model. The findings provide valuable insights for policymakers, industry practitioners, and researchers, facilitating effective management of the lithium supply chain in the face of uncertainty and multiple objectives. The proposed model is distinct in its holistic approach, balancing economic, environmental, and social considerations. It offers a framework for evaluating and optimizing lithium supply chains, taking into account the unique challenges and opportunities presented by the U.S. market. By incorporating stochastic programming and considering regional differences, this research provides a robust tool for decision-makers to navigate the complexities of the lithium industry and drive sustainable growth.

2 - Lithium Recycling: An Experimental Approach

Oluwatosin Atitebi, University of Texas at Arlington, Arlington, TX, United States

As the need for critical minerals like lithium increases for clean energy technologies, new sources will have to be discovered and quickly. However, unlike the fossil fuel energy system the minerals needed for the clean energy system can be recycled. Nonetheless, the science behind recycling lithium is nascent. This work presents our lab scale attempts to efficiently develop experiments to recycle lithium using industrial engineering and operations research techniques.

3 - A Reverse Logistics Supply Chain Network for Sustainable Value Recovery from Li-Ion Batteries in the United States

Apurba Kumar, University of Arizona, Tuscon, AZ, United States

Li-ion batteries (LIB) are increasingly used in renewable energy storage (e.g., solar and wind) and (hybrid) electric vehicles, contributing to the UN sustainable development goals and net zero emissions goal. However, proper management of hazardous wastes after their end of life (EOL) is a great concern. Value recovery from spent LIBs could facilitate sustainable EOL management and recover valuable materials to create a resilient supply of critical raw materials such as lithium, cobalt, and nickel. This study focuses on the design of an optimal reverse logistic supply chain network for sustainable collection and recycling of spent LIBs. An optimization model has been developed to (1) quantify the impact of a governmental policy on promoting the large-scale adoption of LIB recycling technologies and (2) maximize the economic benefits from LIB recycling in the next ten years. The model was applied to the case of United States and suggested the optimal facility locations, processing capacities, recycling technology, and material flows for LIB recyclers. According to our analysis, up to 57% of total available feedstock could be collected over a planning horizon of ten years without any government intervention. However, the government must come forward with properly planned policies to improve the collection rate further. Depending on the budget available to the government, mandating 90% collection of total available spent LIBs and matching 30% of recycling revenue can boost the feedstock collection rate up to 94%.

4 - Critical Minerals for the Clean Energy Transition: An Evaluation

Erick Jones, University of Texas at Arlington, Arlington, TX, United States

Critical minerals, named because of their importance to the economy, are essential for the clean energy economy. The clean energy system is much more mineral intensive than our current fossil fuel based economy. However, like the current economy these minerals are concentrated in locations where securing a supply chain may make it more difficult. This work evaluates modeling strategies to assess this supply chain and shore up the minerals needed for the clean energy transition.

SC77

Regency - 708

Towards a Sustainable Resilient Future - Part B

Invited Session

Computing Society

Chair: Himadri Sen Gupta, University of Oklahoma, 202 W Boyd St, Norman, OK, 73019, United States

Co-Chair: Andrés González, University of Oklahoma, Norman, OK, 73019, United States

1 - A Community Resilience Planning Framework for Maximizing Post-Disaster Healthcare Accessibility

Tasnim Faiz, University of Maryland, College Park, MD, United States, William Hughes, Kenneth Harrison

Accessibility to healthcare services following a disaster is a crucial metric of community resilience. Damages to hospital structures and transportation networks caused by disaster events significantly impact a community's accessibility to these services. Moreover, the increased number of visits to the emergency departments resulting from injured occupants of damaged residential buildings further challenges a community's healthcare system capacity. By adopting suitable mitigation actions, a community can avert the reduction in healthcare accessibility following a disaster, thereby improving resilience. However, identifying suitable mitigation actions for critical infrastructure and buildings, i.e., a community's built environment, to achieve that goal is challenging due to the uncertainty in hazard impact, budgetary limitations, and other restrictions. This work presents a decision-making framework to develop alternative mitigation decision sets for improved healthcare accessibility as a community resilience goal under varied hazard scenarios. The framework adopts a two-stage robust optimization approach, which includes a first-stage model to identify optimal mitigation decisions for maximizing accessibility under a set of hazard impact scenarios and a second-stage model to generate the worst-case scenarios for the current set of actions and evaluate their consequences on a community's built environment and social element. A case study with a community in Shelby County, Tennessee, subjected to earthquake hazards is presented to showcase the applicability of the proposed framework in developing alternative mitigation strategies.

2 - Text Mining of Practical Disaster Reports: Case Study on Cascadia Earthquake Preparedness

Julia Lensing, University of Washington, Lacey, WA, United States

Many practical disaster reports are published every day across the globe in various forms (e.g., after-action reports, response plans, impact assessments, and resiliency plans). These reports enable the next generations to learn from past events to best mitigate and prepare for future disasters. However, the extensive practical literature has limited impacts on research and practice because of the challenge in synthesizing and analyzing the reports. In this study, we 1) present a corpus of practical reports for text mining and 2) an approach to extract insights from the corpus using select text mining tools. We validate the approach through a case study that examines practical reports about the preparedness of the U.S. Pacific Northwest for a magnitude 9 Cascadia Subduction Zone earthquake, which can disrupt lifeline infrastructures for months. The case study illustrated the types of insights that the approach can extract from a corpus. For example, it identified potential differences in priorities between Washington and Oregon state-level emergency management, highlighted latent sentiments expressed in the corpus, and recognized the inconsistent vocabulary across the field. Based on a brief survey of potential user groups, we also discuss opportunities and challenges with text mining of practical disaster reports. For example, simple tools can provide insights potentially only interpretable by those experienced in the field, whereas more complex tools using large language models such as ChatGPT can provide more readily accessible insights, if with known risks of current artificial intelligence. For reproducibility, supporting data and code are made publicly available (DOI: 10.17603/ds2-9s7w-9694).

3 - Machine Learning and Cge Models: A Synergistic Approach to Enhanced Resilience Optimization

Nushra Zannat, The University of Oklahoma, Norman, OK, United States, Charles Nicholson, Harvey Cutler, Himadri Sen Gupta, Andres Gonzalez

This research presents an innovative framework that integrates machine learning (ML) with optimization techniques to analyze and mitigate the impacts of disasters on community infrastructure and the economy. We used penalized regression models to interpret computable general equilibrium (CGE) model outputs and extract predictive insights that quantify the effects of various disaster scenarios as coefficients. The study focuses on utilizing these ML-derived coefficients to refine the objective of an optimization model for enhancing strategic planning in disaster response.

Overall, this research highlights the potential of using advanced analytics in enhancing community resilience -- by transforming data into actionable insights that improve preparedness, response, and mitigation strategies, thereby contributing significantly to sustainable disaster risk management.

4 - Analyzing the Ability of AI Chatbots to Extract Data and Information About Natural Disasters

Ritvik Karthik, Iowa State University, Ames, IA, United States, Cameron MacKenzie, Wilson Diep, Ule Mewanu

Two major trends in the world right now are the increasing capabilities of artificial intelligence (AI) and climate change. One provides great opportunities to advance the world, and the other represents one of the great challenges facing humanity. The risk of natural disasters is increasing due to climate change. Operations researchers frequently sift through a variety of text documents in order to find information and data that can inform a mathematical model for natural disasters. This research examines whether AI chatbots can be used to find accurate data and statistics about natural disasters. The first phase compares the accuracy of ChatGPT 3.5 and Google Bard to answer different questions about natural disasters. The second phase analyzes the ability of ChatGPT 4 to accurately answer questions about a specific natural disaster based on receiving dozens of news articles as an input. In the first phase, Google Bard was more accurate than ChatGPT 3.5 for all the different categories of questions. In the second phase, ChatGPT 4 correctly answered most questions. Our results show there is potential in using AI chatbots to find and reveal information about natural disasters that can be incorporated into operations research models. The ability of AI chatbots to provide input for modelers will likely increase in the future as AI continues to become more capable and accurate.

SC78

Regency - 709

Discrete Optimization and Machine Learning

Invited Session

Computing Society

Chair: Connor Lawless, Cornell University, Ithaca, NY, United States

1 - Physician Rostering with Downstream Capacity Constraints

Yaron Shaposhnik, University of Rochester, Rochester, NY, United States, Yashi Huang, Arik Senderovich

We collaborate with a large hospital that specializes in cancer treatment and develop an interactive interface that optimizes the physicians' roster (i.e., sessions) to improve patient flow and the utilization of resources. We conduct numerical experiments using historical data to assess the accuracy of the infusion load predictions and the potential improvement in different service quality metrics that can be obtained by optimizing the roster.

2 - A Scalable Linear Programming Based Algorithm for Fair K-Means Clustering

Yakun Wang, Lehigh University, Bethlehem, PA, United States, Aida Khajavirad

In [De Rosa and Khajavirad, 2022], the authors introduce a new linear programming (LP) relaxation for K-means clustering, and in [De Rosa and Khajavirad, 2024], they examine its theoretical properties, finding it nearly always tight on various data sets. However, this LP includes $\Omega(n^3)$ inequality constraints, making it impractical for large data sets.

In this work, we propose an efficient algorithm for solving this LP relaxation. The main components of our algorithm are: a separation algorithm to select a proper subset of inequalities, the LP solver *cuPDLP*, which is an efficient GPU-based implementation of the Restarted Primal-Dual Hybrid Gradient method, and warm starting using the effective heuristic *kmeans++*.

We conduct a comprehensive numerical study on real-world data sets from the UCI library.

We are able to solve all instances with up to 2000 data points to global optimality within one hour, hence, significantly outperforming an SDP-based exact solver for K-means clustering.

Subsequently, we consider a variant of fair K-means clustering first proposed by Chierichetti et al. In this framework, each data point is characterized by a sensitive attribute, and a clustering of data is considered fair if the proportion of sensitive attributes in each cluster closely matches their overall proportion in the dataset. We then propose a fair LP relaxation for this problem by showing that clustering fairness can be obtained by adding linear inequalities to the K-means clustering problem.

Our numerical experiments demonstrate the effectiveness of the fair LP relaxation for both synthetic and real-world data sets.

3 - Okridge: Scalable Optimal K-Sparse Ridge Regression

Jiachang Liu, Duke University, Durham, NC, United States, Sam Rosen, Chudi Zhong, Cynthia Rudin

We consider an important problem in scientific discovery, namely identifying sparse governing equations for nonlinear dynamical systems. This involves solving sparse ridge regression problems to provable optimality in order to determine which terms drive the underlying dynamics. We propose a fast algorithm, OKRidge, for sparse ridge regression, using a novel lower bound calculation involving, first, a saddle point formulation, and from there, either solving (i) a linear system or (ii) using an ADMM-based approach, where the proximal operators can be efficiently evaluated by solving another linear system and an isotonic regression problem. We also propose a method to warm-start our solver, which leverages a beam search. Experimentally, our methods attain provable optimality with run times that are orders of magnitude faster than those of the existing MIP formulations solved by the commercial solver Gurobi.

SC99

Flex C

MIF Student Poster Session

Poster Session

Minority Issues Forum

Chair: Karen Hicklin, University of Florida, Gainesville, FL, United States

1 - Optimizing Vehicle Fleet Composition in Underground Mines

John Ayaburi, Colorado School of Mines, Golden, CO, United States, Aaron Swift, Jason Porter, Andrea Brickey, Alexandra Newman

The mining industry relies heavily on the use of diesel-powered equipment, which accounts for heat accumulation and exhaust emissions that can create unsafe working conditions. We present a large-scale production scheduling model that (i) prescribes activity start times at daily fidelity, taking into account ventilation and refrigeration; and, (ii) determines a fleet composition, relative to a diesel-only fleet, that improves productivity. We find that the need for refrigeration is delayed and exhaust emission is reduced as more battery-powered equipment is introduced, showcasing the utility of battery vehicles in maintaining productivity and improving the safety of underground work environments.

2 - TBD

Edikan Udofia, Colorado School of MInes, Golden, CO, United States

TBD

3 - tbd

Sun Ju Lee, Georgia Institute of Technology, Atlanta, GA, United States

tbd

4 - tbd

Rachel Bennett, University of Oklahoma, Oklahoma City, OK, United States

tbd

5 - Race Neutral Calculator Still Disadvantages Black Patients

Amaya McNealey, Georgia Institute of Technology, Atlanta, GA, United States

VBAC-1.0, the vaginal birth after cesarean calculator, incorporated race/ethnicity and faced criticism, leading to the race-neutral VBAC-2.0. Our decision-analytic framework compared both calculators' recommendations, with results showing higher cesarean rates for Black patients under VBAC-2.0 compared to VBAC-1.0.

6 - Distributionally Fair Stochastic Optimization using Wasserstein Distance

Qing Ye, Georgia Institute of Technology, Atlanta, GA, United States

This paper introduces Distributionally Fair Stochastic Optimization (DFSO) based on the Wasserstein fairness measure. DFSO seeks to minimize distributional disparities among groups defined by sensitive attributes, such as gender and race, measured by the Wasserstein distance. We present exact mixed-integer convex programming formulations and propose two lower bounds for the Wasserstein fairness measure. Additionally, we develop efficient algorithms to solve DFSO problems. Numerical studies demonstrate the practical applicability of DFSO in various societally relevant real-world scenarios.

7 - Exploring Nutritional Equity in Foodbank Supply Chains

Nowshin Sharmile, North Carolina A&T State University, Greensboro, NC, United States, Lauren Davis

This research targets enhancing nutritional equity in food distribution by the Foodbank of Central and Eastern North Carolina. It introduces a linear programming model based on the Healthy Eating Research Nutrition Guideline to minimize disparities in access to nutritious food, advancing responses to food insecurity within non-profit operations.

8 - tbd

Shabnam Salehi, The University of Alabama, Tuscaloosa, AL, United States

tbd

9 - tbd

Xiaoquan Gao, Purdue University, West Lafayette, IN, United States

tbd

10 - tbd

Jiannan Xu, Robert H. Smith School of Business, University of Maryland, College Park, MD, United States

tbd

11 - Federated Multiple Tensor-on-Tensor Regression (FedMTOT) for Multimodal Data Under Data-Sharing Constraints

Zihan Zhang, ISyE Georgia Tech, Atlanta, GA, United States, Shancong Mou, Mostafa Reisi, Massimo Pacella, Jianjun Shi

In recent years, diversified measurements reflect the system dynamics from a more comprehensive perspective in system modeling and analysis, such as scalars, waveform signals, images, and structured point clouds. To handle such multimodal structured high-dimensional (SHD) data, combining a large amount of data from multiple sites is necessary (i) to reduce the inherent population bias from a single site and (ii) to increase the model accuracy. However, impeded by data management policies and storage costs, data could not be easily shared or directly exchanged among different sites. Instead of simplifying or facilitating the data query process, we propose a federated multiple tensor-on-tensor regression (FedMTOT) framework to train the individual system model locally using (i) its own data and (ii) data features (not data itself) from other sites. Specifically, federated computation is executed based on alternating direction method of multipliers (ADMM) to satisfy data-sharing requirements, while the individual model at each site can still benefit from feature knowledge from other sites to improve its own model accuracy. Finally, two simulations and two case studies validate the superiority of the proposed FedMTOT framework.

12 - tbd

Jeena Ahuja, Alcuin School, Dallas, TX, United States

tbd

13 - tbd

Himadri Pandey, Georgia Institute of Technology, Atlanta, GA, United States

tbd

14 - tbd

Ogechi Vivian Nwadiaru, University of Massachusetts, Amherst, Amherst, MA, United States

tbd

15 - tbd

Chase Robinson, Colorado School of Mines, Westminster, CO, United States

tbd

16 - Mitigating diesel particulate matter in mine planning

Blaise Yendemeh, Colorado School of Mines, Boulder, CO, United States

tbd

17 - tbd

Raymond Kudzawu-D'Pherdd, Colorado School of Mines, Boulder, CO, United States

tbd

18 - tbd

Felix Ayaburi, Colorado School of Mines, Lakewood, CO, United States

tbd

19 - TBD

Sebastian Rodriguez Cartes, North Carolina State University, Raleigh, NC, United States

tbd

20 - TBD

Saber Dinpazhouh, Rice University, Houston, TX, United States

TBD

21 - Comparative Performance of Battery- and Diesel-Powered Haulage Fleets in Underground Mines

Aaron Swift, Colorado School of Mines, Golden, CO, United States

Battery-electric vehicles can improve ambient conditions in underground mines by reducing the amount of heat and diesel exhaust in working areas. This study uses simulation techniques to determine the effect of electrifying a haulage fleet on the heat load, emissions, and energy usage of an underground mine versus current diesel equipment. This information is important as mine operators assess vehicle electrification as a means to achieve greenhouse gas emissions reduction targets.

22 - tbd

Amy Guo, UC Berkeley, Berkeley, CA, United States

tbd

SD65

Regency - 603

Exploring and Maintaining Teaching Stream Careers in Academia

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, United States

1 - Exploring and Maintaining Teaching Stream Careers in Academia

Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, PA, United States

2 - Panelist

Yaren Bilge Kaya, Columbia University, New York, NY, United States

3 - Panelist

Mariana Escallon-Barrios, Carnegie Mellon University, Pittsburgh, PA, United States

4 - Panelist

Tiffany Bayley, Ivey Business School, London, ON, Canada

Sunday, October 20, 2:15 PM - 3:30 PM

SD01

Summit - 320

New Models and Theory

Contributed Session

Contributed

Chair: Jingchuan Chen, The University of Hong Kong, Hong Kong, N/A, Hong Kong

1 - Scheduling Polices of a Flexible Worker in Two-Station Production Systems

Jingchuan Chen, The University of Hong Kong, Hong Kong, Hong Kong, Shaochong Lin, Zuo-Jun Max Shen

This work is motivated by practical observations in a truck manufacturing line, where some workers are arranged to produce parts for two stations by switching regularly due to the unbalanced productivity of stations. Indeed, the introduction of flexible workers can improve the performance of manufacturing and service systems. Therefore, we study single-cycle scheduling policies for two-stage production systems with one dedicated worker in each station and a flexible worker that can serve both stations. We find that a dedicated allocation policy, i.e., assigning the flexible worker to the station of lower efficiency, is optimal when switches are costly. Besides, our analysis shows that a dynamic scheduling policy is optimal when switch costs are medium, and we provided some instructions for scheduling the flexible worker dynamically. For the system with low-cost switches, we identify the optimal conditions for switching the flexible worker. Finally, we tentatively extend our results for single-cycle scheduling to the policies for multiple cycles.

2 - Minimizing the Stay Time of Open Shop Scheduling Problems with the Unit Processing Time

Gwanhee Lee, POSTECH, Pohang, Korea, Republic of, Kyungduk Moon, Dongyun Kim, Yeonjun Choi, Insoo Park, Juntaek Hong, Kangbok Lee

The open shop scheduling problem involves determining the starting times and routes of jobs across multiple machines. In this study, we introduce a novel service quality metric called "stay time," defined as the duration between a job's starting time and its completion time within a schedule. We examine two objectives related to stay time: the maximum stay time and the total stay time. Our investigation focuses on the computational complexity of scenarios where each job must visit every machine exactly once, with each operation requiring unit processing time, under the condition of an optimal makespan. Additionally, we propose algorithms to address cases that are solvable in polynomial time.

3 - Metric selection and weighted scoring method for determining technology novelty

SungChan Jun, Korea University, Seoul, Korea, Republic of, Chulung Lee

Assessing the characteristics of target technology and determining the value and originality of the technology is crucial for enabling efficient decision-making and creating technology roadmap for countries and companies. While various studies are being conducted to find the value and originality of technology in specific fields using various indicators and classical statistical methodologies, research on a methodology that can analyze the entire industry as a whole is currently insufficient. To address this need, this study developed the following framework: (1) collection of patents related to technologies similar to the target technology, (2) establishment and analysis of technological indicators, (3) determination of weights for each indicator, and (4) evaluation of technology through the derivation of originality scores. The developed framework derives indicators and originality grades based on natural language processing algorithms, ARM algorithms, and network analysis results. Furthermore, this study enhances its credibility by conducting empirical analysis using South Korea's fundamental technology research projects and patents that have been rejected for application. By using the metheodology suggested in this study, the value and originality of the target technology can be converted into objective figures, which will bring great advantages in establishing intellectual property strategies and technology roadmaps.

4 - Predicting promising technologies using a multi-layered network of country, company, and technology

Leehee Kim, Seoul National University, Seoul, Korea, Republic of, Sungjoo Lee

Identifying promising technology is crucial for companies and countries to establish effective technology strategies. and accordingly, much effort has been made to develop an approach to predict such technologies. Despite such effort, however, few studies considered the intertwined effect of various factors on technology growth. To fill the research gap, this study aims to propose a patent-based approach to predict promising technologies using a multi-layered network consisting of three layers – country, company, and technology. To achieve this aim, firstly, similar patents are clustered to define technologies, with each cluster named according to the technology area of the patent. A heterogeneous multilayer network is then constructed to illustrate the relationships between, countries, companies, and technologies, with each entity serving as a node. Promising technology clusters are predicted using indicators from three perspectives: country, company, and technology . Following this prediction, the research findings are expected to provide valuable insights for technology strategies, identifying the country-company-technology community through multilayer overlapping community detection and analyzing the impact of each country and company on the emergence of specific technology as promising fields. Ultimately, this study seeks to elucidate how countries and companies influence the development of specific technology into promising fields. The results are expected to assist companies and countries in formulating technology strategies by identifying promising technology clusters within their areas of interest.

5 - An Integrative Framework for The Sustainable Supply Chains; From Operational to ESG-Based Performance Metrics

Leili Soltanisehat, University of Massachusetts Dartmouth, Boston, MA, United States

Aligned with the Sustainable Development Goals (SDGs), companies are increasingly integrating Environmental, Social, and Governance (ESG) considerations into their investment strategies and operations management. Supply chain risk management is one of the most important targets of SDGs, as a sustainable supply chain surpasses the conventional emphasis on cost-cutting and efficiency. Through the adoption of sustainable supply chain practices, organizations can not only improve their environmental footprint but also attain sustainable financial and operational viability. However traditional supply chain risk assessment tools often fall short of comprehensively addressing the multifaceted challenges that modern supply chains face. Traditional tools often focus on narrow aspects of risk, such as operational disruptions or supplier reliability, neglecting broader environmental, social, and governance (ESG) factors. In addition, traditional tools may not facilitate collaboration and information sharing across supply chain partners (e.g., echelons), limiting their ability to identify and mitigate the propagated and accumulated risks. Therefore, to cover this gap, this research aims to design a universal supply chain risk assessment tool, that utilizes a mathematical modeling approach to integrate supply chain network dynamics, the material flows, and the risk propagation within the supply chain network. The proposed framework incorporates various aspects of the risks affecting the performance of the supply chain including, ESG risks, operational risks, and material risks. In addition, the risk assessment tool provides multiple scales of the risk measurements such as financial (profit and cost), time (delivery time), and social (affected community) scales.

SD02

Summit - 321

IISE Transactions Invited Session

Award Session

Quality, Statistics and Reliability

Chair: Jionghua Jin, University of Michigan, Ann Arbor, MI, United States

1 - Dynamic exploration–exploitation trade-off in active learning regression with Bayesian hierarchical modeling

Ashif Iquebal, Arizona State University, Tempe, AZ, United States

Active learning provides a framework to adaptively query the most informative experiments towards learning an unknown black-box function. Various approaches of active learning have been proposed in the literature, however, they either focus on exploration or exploitation in the design space. Methods that do consider exploration–exploitation simultaneously employ fixed or ad-hoc measures to control the trade-off that may not be optimal. In this article, we develop a Bayesian hierarchical approach, referred to as BHEEM, to dynamically balance the exploration-exploitation trade-off as more data points are queried. To sample from the posterior distribution of the trade-off parameter, we subsequently formulate an approximate Bayesian computation approach based on the linear dependence of queried data in the feature space. Simulated and real-world examples show the proposed approach achieves at least 21% and 11% average improvement when compared to pure exploration and exploitation strategies, respectively. More importantly, we note that by optimally balancing the trade-off between exploration and exploitation, BHEEM performs better or at least as well as either pure exploration or pure exploitation.

2 - Simultaneous optimal control of directional missed discovery rates in data stream diagnosis

High-dimensional data streams are ubiquitous in modern manufacturing, due to their ability to provide valuable information about the industrial system's performance on a real-time basis. If a shift occurs in a production process, fault diagnosis based on the data streams is of critical importance for identifying the root cause. Existing methods have largely focused on controlling the total missed discovery rate without distinguishing missed signals for positive versus negative components of the shift vector. In practice, however, losses incurred from the two directional shifts can differ substantially, so it is desirable to constrain the proportions of missed signals for positive and negative components at two distinctive levels. In this article, we propose a fault classification procedure that controls the two proportions separately. By formulating the problem as Lagrangian multiplier optimization, we show that the proposed procedure is optimal in the sense that it minimizes the expected number of false discoveries. We also suggest an iterative adjustment algorithm that converges to the optimal Lagrangian parameters. The asymptotic optimality for the data-driven version of our procedure is also established. Theoretical justification and numerical comparison with state-of-the-art methods show that the proposed procedure works well in applications.

3 - 0 CrossRef citations to date 0 Altmetric Data Science, Quality & Reliability Inspection policy optimization for hierarchical multistate systems under uncertain mission scenarios: A risk-averse perspective

Most engineered systems intend to perform missions with a pre-specified target success probability to reduce undesirable failure risks. Before executing the next mission, inspection activities are conducted across various physical levels for assessing the probability of mission success. However, due to the randomness of a system's degradation behavior and the presence of measurement errors, inspection results inevitably contain uncertainty. Meanwhile, mission durations and acceptable system states may also be uncertain, due to uncontrollable factors, such as random operating environments and mission demands. In such a circumstance, it is of great significance to identify the optimal multilevel inspection policy to answer, as great confident as possible, the question that the system can complete the next mission with a target mission success probability. This paper develops a novel metric to gauge the effectiveness of a multilevel inspection policy to assess if the system can complete the next mission with a pre-specified target success probability from a risk-averse perspective, based on which an optimization method is proposed to seek an inspection policy under uncertain scenarios with the aim of minimizing the maximum regret of the proposed metric. A stochastic fractal search algorithm, along with two tailored local search rules, is designed to efficiently resolve the resulting optimization problem. Two cases, including a three-component system and a rocket fueling mechanism's control system, are used to illustrate the efficacy of the proposed approach, which is capable of effectively identifying the risk of mission failures by inspection policies.

SD03

Summit - 322

Deep Neural Network Methodologies for High Dimensional Data in Engineering Applications

Invited Session

Quality, Statistics and Reliability

Chair: Wenbo Sun, University of Michigan Transportation Research Institute, Ann Arbor, MI, United States

Co-Chair: Bo Shen, New Jersey Institute of Technology, Newark, NJ, United States

1 - A Deep-Learning-Based Solution for Interpretable in-Situ Monitoring in Additive Manufacturing

Xiaoyang Song, University of Michigan, Ann Arbor, MI, United States, Wenbo Sun, Metin Kayitmazbatir, Judy Jin

In situ quality inspection is a key challenge in additive manufacturing. In this work, we developed a deep-learning-based model to monitor the product porosity using purely optical emission spectra, which has been proven useful in detecting microscopic pores. Although recurrent neural networks and transformers exhibit decent performance in sequential data processing, feature extraction from high-dimensional noisy spectra is still a challenging task. To overcome the challenges, we also collect layer-wise CT scan data after the manufacturing process and use them to guide the learning of spectra features. Specifically, we build connections between the spectra and CT scans using deep canonical correlation analysis to ensure that the learned features are good representatives with those defects, which are presented in the CT scans. We experiment with the proposed method on both simulation data and data that are collected during the real manufacturing process. In addition, as the structure of the proposed multimodal model is akin to that of the real data generation process, it also exhibits decent interpretability compared to other deep-learning-based methods.

2 - Active Multi-Modal Data Fusion for Physiological Signals

Yue Zhao, Rensselaer Polytechnic Institute, Tory, NY, United States, Yinan Wang

Exploring human cognitive states through multi-modal physiological signals from multiple sources by using machine learning method is increasingly significant and has attracted considerable attention in recent research. Machine learning methods often require large datasets, but extensive data collection may be limited due to the significant costs and time investment. Multi-modal data fusion integrates data from multiple sources and is supposed to provide a more comprehensive understanding. However, fusing multi-modal data can usually include misleading information due to noise and data quality issues. Additionally, multi-modality datasets often contain redundant information, so it is essential to identify a subset of modalities with complementary information along with high-quality and informative data samples to improve the performance of the ML method. There are two main challenges: (1) which combination of modalities will benefit the downstream task the most and (2) how to evaluate the informativeness and quality of data samples. We proposed the active multi-modal data fusion framework by integrating Active Learning and Bayesian Optimization. Our proposed framework is validated on collected multi-modal physiological data from the pilots who were tested by the simulated Instrumented Landing System (ILS). The objective is to classify the level of difficulty of flight conditions using the multi-modal physiological data. The result indicates that our framework can select the combination of modalities with complementary information and high-quality data samples to improve performance while has a better sample efficiency.

3 - Transformer-Empowered Incremental Learning for Preventive Defect Monitoring in Additive Manufacturing

Boris Oskolkov, Oklahoma State University, Stillwater, OK, United States, Zhangyue Shi, Chenang Liu

Additive Manufacturing (AM) enables unparalleled design freedom and customization capabilities in modern manufacturing. However, the occurrence of defects during the AM process remains a significant challenge, hindering wider adoption and limiting the reliability of AM-produced parts. However, traditional methods relying on computationally expensive simulations or time-consuming post-build inspections are often inadequate for the dynamic nature of AM processes. Our work introduces an improved approach for preventive defect monitoring in AM that leverages the power of transformers and incremental learning. Transformers, renowned for their ability to capture complex temporal short-term and long-term relationships in sequential data, are employed to learn intricate patterns from in-situ sensor data collected during the AM process. This enables the model to predict defects before they manifest, facilitating proactive intervention and mitigation strategies. The transformer is accompanying our previously introduced incremental learning framework which empowers the model to continuously evolve and refine its predictive capabilities as the process environment evolves. We will showcase experimental results demonstrating the effectiveness of our approach in real-world AM scenarios. Specifically, we will highlight the model's ability to accurately predict defects in real-time, its adaptability to evolving AM environments, and its potential to significantly enhance the quality and efficiency of AM production within cost-related manufacturing constraints.

4 - Detecting Geometric Defects in Gears Using Deep Learning of 3D Point Clouds

Alice Mei, University of Michigan Ann Arbor, Ann Arbor, MI, United States

Geometric integrity is vital in manufacturing as it affects a product’s functionality, reliability, and safety, making accurate geometric assessments critical for quality control. The recent advancement of three-dimensional (3D) metrology has greatly improved the precision of examining geometric integrity, focusing on dimensional accuracy, surface quality, and shape conformity. However, the high dimensionality, unstructured data, and sparseness in defective regions present significant challenges in utilizing high-resolution 3D metrology for advanced manufacturing processes.

To address these issues, our study proposes the "MFGNet-gear" dataset, a scalable and specialized manufacturing dataset featuring twelve part designs, each offered in four quality levels. We have adapted the PointNet++ framework to develop a deep learning model that efficiently performs end-to-end analysis of 3D point clouds. This model is designed to carry out various analytical tasks, which include the classification of part designs and detection of a range of geometric defects. Applied to the "MFGNet-gear" dataset, the model achieved up to 100% accuracy in gear design classification and 85% in discerning different quality levels. Through a series of experiments, we’ve assessed how changes in measurement resolution and precision influence model performance. Results obtained also highlight the potential of deep learning in automating 3D point cloud analysis for quality control applications beyond gear manufacturing. Lastly, our work suggests potential future research directions, including the development of new deep learning architectures specifically designed for manufacturing 3D point clouds and strategies for adaptive measurement planning.

SD04

Summit - 323

Advanced Machine Learning and Artificial Intelligence for Healthcare Diagnostics

Invited Session

Quality, Statistics and Reliability

Chair: Ashif Iquebal, Arizona State University, Tempe, AZ, United States

Co-Chair: Zimo Wang, SUNY Binghamton, Binghamton, NY, United States

1 - Long term prediction of dynamic intermittency with application for telecardiology based monitoring and prognosis

Zimo Wang, SUNY Binghamton, Binghamton, NY, United States, Qiyang Ma

The performance of long-term prediction models is currently impeded due to the mismatch between the nonstationary representations of statistical learning models and the underlying dynamics from real-world systems, which results in low long-term prediction accuracies for many real-world applications. We present a Recurrent Gated Unit-based Mixture Kriging Machine Bayesian Filtering (ReGU-MKMBF) approach for characterizing nonstationary and nonlinear behaviors of one ubiquitous real-world process—dynamic intermittency. It models the transient dynamics in the state space as recurrent transitions between localized stationary segments/attractors. Then, a case study on predicting the onset of pathological symptoms associated with Electrocardiogram signals is presented. The results suggest that ReGU-MKMBF improves forecasting performance by extending the prediction time horizon by an order of magnitude while maintaining high accuracies for the foreseen estimates. Implementing the presented approach can subsequently change the current scheme of online monitoring and aftermath mitigation into a prediction and timely prevention for telecardiology.

2 - Talk1

Zeynep Ertem, SUNY Binghamton, Vestal, NY, United States

tbd

3 - Counterfactual explanations to generate clinical inference into genetic mutation prediction from black-box models

Ishraq Rahman, Arizona State University, Tempe, AZ, United States, Ashif Iquebal

We investigate the problem of predicting and explaining genetic mutation in breast cancer patients derived from high-dimensional magnetic resonance imaging (MRI) without any tumor segmentations. We propose a sparse counterGAN that leverages a generative adversarial model to generate counterfactual MRI instances by learning the true data distribution. During the training process, the generator network generates residuals that are added to the original MRI instance to produce a counterfactual instance. The discriminator network tries to distinguish the counterfactual instance from the original instance. A pre-trained classifier network then classifies the original and counterfactual examples into different categories, such as “mutation” and “non-mutation”. We regularize the loss function to include a p-norm regularization. With p-norm regularization, we can set the value of p to control the sparsity of the solution, where a larger value of p corresponds to a sparser solution with fewer changed features. Finally, we present our results obtained from clinical MRI data as well as inference from domain experts in explaining the genetic prediction results.

SD05

Summit - 324

DEI Best Student Paper Award I

Award Session

Diversity, Equity, and Inclusion

Chair: Zeyu Liu, West Virginia University, Morgantown, WV, United States

Co-Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Designing Size Inclusive Fashion Assortments

Alireza Alavi, McGill University, Montreal, QC, Canada, Mehmet Gumus, Javad Nasiry

In fashion retailing, size unavailability occurs when displayed items are unavailable in some sizes, leading to lost sales. Plus-size customers encounter size unavailability far more frequently than others, contributing to a "subtle form of social discrimination" that frustrates customers and harms brand perception. This raises a critical question: How should an online fashion retailer design its assortment to be both size inclusive and profitable? In this work, we answer this question by developing a choice model that incorporates customers' reactions to size unavailability, framing the problem as a zero-sum game between the retailer and an adversary (nature), and designing (slow) exact and (fast) approximation algorithms to find the equilibrium of this game. We calibrate our model using data from a large European-based fashion retailer. Our simulations show that our approximation algorithm can reduce unfairness by 30% with only a 1.5% reduction in revenue.

2 - Pay Transparency in Heterogeneous Teams: Could Some Fairness Concerns Hurt Productivity?

Guillaume Roels, INSEAD, Fontainebleau, France, Lin Chen

In homogeneous teams, pay transparency, which induces agents to compare themselves to others, leads to a higher output, results in no inequalities in equilibrium, and is inclusive (i.e., all agents are engaged). This is true irrespective of the basis of social comparisons: income, utility, or reward. How about heterogeneous teams? Could pay transparency backfire under some types of fairness concerns? We consider a model of a principal and two agents of different abilities who are averse to inequalities between each other under pay transparency. We study all three bases of comparison. For each type of fairness concern, we characterize when pay transparency leads to a higher or lower payoff for the principal, results or not in inequalities between agents, and involves only one or both agents in production. We find that income fairness never benefits the principal and might be perceived as unfair and non-inclusive, in contrast to reward fairness, which is always beneficial on all three fronts. Utility fairness lies somewhere in between, offering benefits on all three fronts when agents are not so heterogeneous, but hurting fairness and inclusivity without changing the principal's payoff otherwise. Although pay transparency can help close the gender pay gap, anchoring agents to income fairness by disclosing only their incomes can hurt their productivity while creating feelings of unfairness and lack of inclusivity. In contrast, anchoring them to reward or utility fairness by contextualizing their incomes with their individual contributions or costs may be beneficial on all three fronts.

3 - From Black to Grey: Improving Access to Antimalarial Drugs in the Presence of Deceptive Counterfeits

Jiatao Ding, Nanyang Technological University, Singapore, Singapore, Michael Freeman, Sasa Zorc

In malaria-endemic countries, limited access to affordable antimalarial drugs has fueled the proliferation of counterfeit medications. This paper examines such markets to determine how philanthropic donors can optimize fund allocation to subsidize antimalarial drug distribution through private-sector channels. We develop a game-theoretic model where retailers can source drugs from certified suppliers, uncertified (potential counterfeit) suppliers, or both. Unlike previous studies, we show that relying solely on purchase subsidies may not be optimal when counterfeits are present. Under donor budget constraints, sales subsidies that account for both legitimate and counterfeit drugs may be more effective, or it may be better to avoid subsidies altogether. We also analyze five non-subsidy strategies — improving consumer awareness, penalizing counterfeit sourcing, adopting traceability technology, cracking down on counterfeit supplies, and enforcing price controls — to determine when these can enhance or hinder outcomes. A numerical analysis, calibrated to Mozambique’s malaria data, provides further practical insights. Our findings emphasize the importance of tailoring subsidy schemes and policy choices to market conditions, such as retailer pricing power and demand uncertainty, to improve access to life-saving medicines. These insights could be applied to other endemic diseases, offering a broader framework for enhancing public health in resource-constrained settings.

SD06

Summit - 325

Emerging Topics in Location Analysis

Invited Session

Location Analysis

Chair: Gina Galindo, Universidad del Norte, Puerto Colombia, Colombia

1 - *Developing a Statewide Active Transportation Planning Dashboard with Mobility Data*

Rebecca Bian, Louisiana Transportation Research Center; Louisiana State University, Baton Rouge, LA, United States

Existing walking/biking data are often insufficient to support active transportation planning activities. This study used an emerging large-scale human mobility dataset to identify places where there are a higher number of short-distance trips to public places (expressed by a mobility index). Demographic variables were integrated into the mobility index to prioritize access for more people and address equity concerns. In addition, a safety index (reflecting crash injuries/fatalities involving pedestrians/bicyclists) and a connectivity index (reflecting the density of existing active transportation infrastructure) were also considered together with the mobility index in generating an investment potential score. All the indices/scores were mapped at hexagon and segment level on an online dashboard. The developed dashboard is expected to support statewide active transportation planning and project selection/prioritization decision-makings. The proposed methodology is replicable for any places that have few active transportation facilities and where pedestrian/bicyclist count data are not sufficient to model demand directly.

2 - Powering routes: Strategically locating chargers for electric trucks' viability

Carlos Otero, Universidad Del Norte, Barranquilla, Colombia

Besides their high procurement prices, battery electric heavy-duty trucks are also limited in range, which is one of the biggest barriers to their adoption. It is particularly difficult to deal with this in the for-hire trucking market, where demand is highly dynamic and geographically dispersed. Despite the lower operative cost of electric trucks compared to diesel trucks, the limited income due to serving a smaller number of loads does not offset profit. An alternative solution is to strategically install electric chargers along freight corridors. According to recent research, locating charging stations at different distances may improve the profitability of these trucks. Here, it is presented the results of a study based on real-life data, which combines the simulation of cargo generation and the routing of fleets of electric and diesel trucks. Results indicate that optimizing fleet composition, charging station locations, and routing decisions may lead to profitable trucking operations in favor of the adoption of cleaner vehicles.

3 - Strategic Evacuation Planning: Integrating Road Network Performance Risk for Natural Hazard Preparedness

Daniel Rivera, University of California Davis, Davis, CA, United States, Miguel Jaller

In this paper, the authors devise a method to design an evacuation plan for preparing for natural hazards such as wildfires, flash floods, and other hazards that require evacuations. The authors examine the critical aspects of the road network infrastructure and their influence on estimating the evacuation performance (EP). They use the Road Network Performance (RNP) Risk to prioritize the evacuation process for the most vulnerable individuals in a case study in California. The purpose of this paper is to determine the minimum evacuation time in a staggered evacuation, where the evacuation will be conducted using the minimum set of evacuation paths at the minimum risk cost. The authors will combine a set of mathematical programming optimization models with RNP risk based on historical data of natural hazards, socio-economic characteristics of the population, and topological characteristics of the road network. This methodology will provide the evacuation scheduling (traffic generation and distribution) and the routes (traffic assignment) that optimize the time, risk, and utilization of resources (paths). The authors demonstrate how the evacuation time and total risk perceived by evacuees correspond to the socio-economic characteristics of the population. Additionally, a set of scenarios are conducted to determine the sensitivity of the evacuation plans in terms of the evacuation time and the overall perceived risk in a set of cities in California. Finally, the authors provide policy insights for the development of evacuation plans.

SD07

Summit - 327

Language Models and Intelligible AI

Flash Session

Contributed

Chair: Tyler Sam, Cornell University, Ithaca, NY, United States

1 - The Limits of Transfer Reinforcement Learning with Latent Low-rank Structure

Tyler Sam, Cornell University, Ithaca, NY, United States, Yudong Chen, Christina Yu

While reinforcement learning algorithms are increasing in popularity, many algorithms are too costly due to the size of the problem's state space S or action space A. To resolve this issue, we study transfer reinforcement learning with latent low rank structure, where the learner interacts with M source MDPs and attempts to learn a latent representation of the action, states, or both to use in the target MDP to greatly improve performance. Specifically, we study the settings in which the source and target MDPs have transition kernels with Tucker rank $(|S|, d, |A|)$, $(|S|, |S|, d), (d, |S|, |A|)$, or $(|S|, d, d)$. In these settings, our algorithm first learns a near-optimal Q function or estimate of transition kernel for each of the source MDPs and constructs a latent feature representation with the singular vectors of the estimate Q function. Then, our algorithm employs a modification of LSVI-UCB, a common linear MDP algorithm \cite{Jin2020ProvablyER}, to remove the dependence on $|S|, |A|, $ or both in the regret bound in the target MDP. Furthermore, we introduce the transfer-ability coefficient that quantifies the difficulty in performing transfer learning on a specific problem. We then prove lower bounds that show our algorithms are optimal with to the transfer-ability coefficient.

2 - ChatGPT and Firms Performance

Aida Sanatizadeh, Northern Illinois University, Dekalb, IL, United States, Hamid Vakilzadeh, David Wood

This abstract investigates the impact of Large Language Models (LLMs) like ChatGPT on the quality of corporate earnings announcements (EAs). Utilizing AI tools enhances the precision and clarity of financial disclosures, potentially influencing investor decisions and market efficiency. Our study uses data from S&P500 non-English companies, employing event study and difference-in-differences methodologies to assess changes in EA characteristics post-ChatGPT. This research contributes to understanding how generative AI reshapes financial communication and market dynamics.

3 - Pandemic Effect on Online Reviews: The Role of Restrictive Policies, Elite Users, and Political Leanings

Xilu Zhai, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Jifeng Luo

This study investigates the impacts of government-imposed restrictive policies on consumer behavior and online sentiment within the restaurant industry during the COVID-19 pandemic. Utilizing Yelp review data and the Oxford COVID-19 Government Response Tracker, the research employs a continuous difference-in-differences methodology to analyze how early review sentiments and the strictness of government policies influence online review sentiments and ratings. Key findings suggest that positive emotions in early reviews amplify the negative impact of the pandemic on subsequent user ratings. Moreover, stringent government policies, while necessary for public health, significantly exacerbate public unrest and negatively affect online review sentiments. Political orientations also play a crucial role, with consumers in liberal and swing states showing heightened sensitivity to policies compared to those in conservative states. These insights are vital for policymakers and businesses in crafting strategies that balance public health priorities with minimizing adverse psychological impacts during health crises.

4 - Self-Interest and Systemic Benefits: Emergence of Collective Rationality in Mixed Autonomy Traffic Through Deep Reinforcement Learning

Di Chen, University of California, Davis, Davis, CA, United States, Jia Li, Michael Zhang

Autonomous vehicles (AVs) are expected to be commercially available in the near future, leading to mixed traffic with human-driven vehicles (HVs). Although studies have shown that AVs can be deployed to benefit the overall traffic system performance by incorporating system-level goals into their decision making, it is not clear whether they remain beneficial when acting out of self-interest—a trait common to all driving agents, both human and autonomous. This study aims to explore the potential of leveraging self-interested AVs to bring benefits to all driving agents in mixed autonomy traffic systems. We introduce the concept of collective rationality (CR) in mixed traffic and examine its properties in micro-simulation. This concept, originating from game theory and behavioral economics, means that driving agents may form cooperation at collective level even when pursuing individual interests. Our recent research has proven the existence of CR in an analytical game-theoretical model and empirically in mixed human-driven traffic. In this paper, we demonstrate CR can be attained among driving agents trained with deep-reinforcement learning and a simple reward design. We examine the extent to which self-interested traffic agents can achieve CR without directly incorporating system-level objectives. Results show the consistent emergence of CR, which indicates the robustness of these properties. We also postulated and analyzed the underlying mechanism to explain the emergence of CR in the microscopic, dynamic environment. These findings paved the way for the adoption of advanced learning methods, such as federated learning, to achieve cooperation spontaneously among self-interested driving agents in mixed-autonomy systems.

5 - The Individual&Rsquo;S Perception of AI and Its Effects on The Organization&Rsquo;S Ai Initiatives

Catherine Petersen, Johnson & Wales University, Mont Vernon, NH, United States

Investment in Artificial Intelligence (AI) does not result in returns for business despite improving technology. This study investigated one potential reason for the failure of AI when implemented within an organization with the intent to drive success with this technology. AI is unique amongst technologies due to its ability to inspire emotions from potential users and these emotions can affect the individual’s willingness to adopt the technology.

Using an explanatory sequential research design, this study examined the relationship between how an individual perceives AI and their willingness to accept the technology in their role within an organization. It also considered the role of experience and an individual’s perception of AI.

The results found that including AI in the description of the initiative decreased support from 81.5% to 54.8%. There was also a significant relationship between an individual’s perception of AI and their support of an initiative that specifically names AI ($p < .05$). In interviews, nine out of ten respondents discussed job replacement and showed a tension between hopes and fears with regard to the technology. No statistically significant relationship was found between an individual’s experience with AI and their perception of AI but there was a relationship between experience and their expectations for effort and performance. Interviews showed that an individual’s perception of job replacement or challenges with the technology were not affected or increased with experience using AI. Based on these findings, managers can better determine if their organizations are ready to implement AI successfully.

SD08

Summit - 328

OR in Security & Defense III (Mission Support)

Invited Session

Military and Security

Chair: Kyle Swan Hovey, Department of Defense, Columbia, MD, United States

1 - Modeling Attrition for More Accurate Hiring Goals

Beth Goodman, DoD, Laurel, MD, United States, Alex Schulte, Sarah Charlton

A challenging problem for any large organization is ensuring that the workforce is adequate to meet the requirements of that entity. Understanding who and how many people are leaving (attrition) can assist with planning for replacements, and therefore inform hiring goals. We have used a variety of forecasting techniques and historical data to select models that can leverage trends to suggest a more robust estimate of overall (and a selection of subcategories) attrition.

2 - UAS Swarms and Swarm Intelligence: The Current State and Future Concepts

Cameron Carter, USAF, Baltimore, MD, United States

The adoption of unmanned systems in the US military has been a disruptive shift in the realm of modern warfare. Unmanned Aircraft Systems (UAS) Swarms are the next iteration in the shift to drone technology, particularly as the development of Swarm Intelligence and better autonomous systems accelerates. Our efforts aim to examine the current literature and body of research on Swarm Intelligence and the types of AI models and algorithms fueling it. We will identify the current levels of autonomy and human involvement in relation to drone swarms, the current technological limitations, and where we believe the capabilities will be in the future.

This research will also examine the current efforts from USG agencies as well as military research laboratories in order to capture the current state of UAS swarm testing and development and their vision for the future of swarm technology. We will explore the defense applications and benefits of these low-equity, unmanned systems highlighting the effectiveness and adaptability that comes with their dynamic approach. Our efforts will take a look at the legal and ethical considerations surrounding the use of intelligent UAS swarms for both offensive and defensive applications. We aim to provide a comprehensive analysis of the current and potential future capabilities of UAS swarms in the context of a military application.

3 - Enhancing Textual Data Classification with Advanced Language Models: An Evaluation of GPT, BARD, and LLAMA

andrew gelbard, Department of Defense, Annapolis Junction, MD, United States

The accurate classification of textual data is critical for many organizations, but it can be a challenging and time-consuming task that depends on contextual interpretations. Our research aims to explore the potential of advanced language models, such as GPT, BARD, and LLAMA, to automate and refine the task of text classification. We will fine-tune these foundation models to assess their efficacy in distinguishing between different types of textual data. This approach could dramatically enhance decision-making speeds, reduce human workload, and ensure consistent information security across various organizations. We will use a dataset of textual data to test the adaptability and accuracy of these AI models, and we expect that our results will have a broader applications in automatic data handling and security protocols in various sectors.

4 - Internship Opportunities within the DoD

Kyle Hovey, Department of Defense, Columbia, MD, United States

The U.S. Department of Defense has a long history of employing Operations Researchers to tackle some of its most complex problems. To ensure the DoD continues to bring in exceptional operations researchers, there are a number of well-established internship opportunities within the DoD targeting those pursuing degrees in Operations Research and related fields. In this talk, we will discuss several of these opportunities and provide additional information on how to apply. Note: while this talk is open to any interested participants, most DoD internships are restricted to U.S. citizens only.

SD09

Summit - 329

Michael H. Rothkopf Junior Researcher Paper Prize

Award Session

Auctions and Market Design

Chair: Sasa Pekec, Duke University, Durham, NC, United States

1 - Pricing Optimal Outcomes in Coupled and Nonconvex Markets: Theory and Applications to Electricity Markets

Johannes Knoerr, Technical University of Munich, Munich, Germany, Martin Bichler, Mete Ahunbay

coming soon

2 - Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances

Akshit Kumar, Columbia Business School, New York, NY, United States, Omar Besbes, Yash Kanoria

coming soon

3 - Dynamic Matching with Post-Allocation Service and its Application to Refugee Resettlement

Soonbong Lee, Yale University, New Haven, CT, United States, Kirk Bansak, Vahideh Manshadi, Rad Niazadeh, Elisabeth Paulson

coming soon

SD10

Summit - 330

Prompt Engineering, Fine Tuning, Alignment and Market Applications of Large Language Models

Invited Session

Auctions and Market Design

Chair: Wanning Chen, University of Washington, Seattle, WA, United States

1 - Risk Control for Large Language Models

Catherine Chen, Stanford University, Stanford, CA, United States, Lihua Lei

Recent developments in large language models (LLMs) have led to their widespread usage for various tasks. The prevalence of LLMs in society implores the assurance on the reliability of their performance. In particular, risk-sensitive applications demand meticulous attention to unexpectedly poor outcomes, e.g., toxic answers, humiliating language, and offensive outputs. In this work, we present an approach that is guaranteed to control any distortion risk measure, characterized by a weighted average of quantiles of the loss incurred by the LLM, without making any assumptions on the statistical properties of the LLM. Our method exploits the connection between conformal risk control and L-statistics, a type of estimator that dates back to the 40s, and leverages asymptotic properties of L-statistics.

2 - Rlhf and Independence of Irrelevant Alternatives: Perverse Incentives

Wanqiao Xu, Stanford University, Stanford, CA, United States, Shi Dong, Xiuyuan Lu, Grace Lam, Zheng Wen, Benjamin Van Roy

Existing algorithms for reinforcement learning from human feedback (RLHF) can incentivize responses at odds with preferences because they are based on models that assume independence of irrelevant alternatives (IIA).
The perverse incentives induced by IIA hinder innovations on query formats and learning algorithms.

3 - A Probabilistic Approach for Model Alignment with Human Comparisons

Junyu Cao, The University of Texas at Austin, Austin, TX, United States, Mohsen Bayati

A growing trend involves integrating human knowledge into learning frameworks, leveraging subtle human feedback to refine AI models. Despite these advances, no comprehensive theoretical framework describing the specific conditions under which human comparisons improve the traditional supervised fine-tuning process has been developed. To bridge this gap, this paper studies the effective use of human comparisons to address limitations arising from noisy data and high-dimensional models. We propose a two-stage "Supervised Fine Tuning+Human Comparison" (SFT+HC) framework connecting machine learning with human feedback through a probabilistic bisection approach. The two-stage framework first learns low-dimensional representations from noisy-labeled data via an SFT procedure, and then uses human comparisons to improve the model alignment. To examine the efficacy of the alignment phase, we introduce a novel concept termed the "label-noise-to-comparison-accuracy" (LNCA) ratio. This paper theoretically identifies the conditions under which the "SFT+HC" framework outperforms pure SFT approach, leveraging this ratio to highlight the advantage of incorporating human evaluators in reducing sample complexity. We validate that the proposed conditions for the LNCA ratio are met in a case study conducted via an Amazon Mechanical Turk experiment.

4 - Orchestrating Language Models for Supply Chain Optimization

Beibin Li, Microsoft Research, Redmond, WA, United States, Konstantina Mellou, Luke Marshall, Ishai Menache

Supply chain operations traditionally involve a variety of complex decision making problems. Over the last few decades, supply chains greatly benefited from advances in computation, which allowed the transition from manual processing to automation and cost-effective optimization. Nonetheless, business operators still need to spend substantial efforts in explaining and interpreting the optimization outcomes to stakeholders. Motivated by the recent advances in Large Language Models (LLMs), we study how this disruptive technology can help bridge the gap between supply chain automation and human comprehension and trust thereof. We design OptiGuide -- a framework that accepts as input queries in plain text, and outputs insights about the underlying optimization outcomes. Our framework does not forgo the state-of-the-art combinatorial optimization technology, but rather leverages it to quantitatively answer what-if scenarios (e.g., how would the cost change if we used supplier B instead of supplier A for a given demand?). Importantly, our design does not require sending proprietary data over to LLMs, which can be a privacy concern in some circumstances. We demonstrate the effectiveness of our framework on a real server placement scenario within Microsoft's cloud supply chain. Along the way, we develop a general evaluation benchmark, which can be used to evaluate the accuracy of the LLM output in other scenarios.

SD11

Summit - 331

Network Analysis and Graph Neural Networks II

Invited Session

Telecommunications and Network Analytics

Chair: Wenting Li, Los Alamos National Laboratory, Los Alamos, NM, United States

Co-Chair: Lili Zhang, Hewlett Packard Enterprise, Suwanee, GA, United States

1 - Empower Graph Machine Learning for Trustworthy Decision Making

Jundong Li, University of Virginia, Charlottesville , VA, United States

Graph machine learning (GML) models, such as graph neural networks, have proven to be highly effective in modeling graph-structured data and achieving remarkable predictive performance in various high-stake applications, including credit risk scoring, crime prediction, and medical diagnosis. However, concerns have been raised regarding the trustworthiness of GML models in decision making scenarios when fairness, transparency, and accountability are lacking.

To address these concerns, I will present our recent work on empowering GML for trustworthy decision making by focusing on three key aspects: fairness, explanation, and causality. First, I will discuss how to improve the fairness of GML from a data debiasing perspective. In particular, I will show how to measure data biases regarding different modalities of graph data and how to mitigate the data biases in a model-agnostic manner that can benefit different GML models. Second, I will show that explanation, as an effective debugging tool, not only can help us understand how the decisions are made but also could serve as a useful tool to diagnose how biases and discrimination are introduced in GML. Toward this goal, I will present a post-hoc structural explanation framework that can understand the unfairness issues of GML.

Third, I will argue the emerging need to introduce causality for trustworthy decision making on graphs, as traditional GML could heavily rely on spurious correlations for making decisions. To bridge the gap, I will present a GML-based causal inference framework that aims to unleash the power of graph information for causal effect estimation.

2 - Signal Processing over Multilayer Graphs: Theoretical Foundations and Practical Applications

Songyang Zhang, University of Louisiana at Lafayette, Lafayette, LA, United States, Qinwen Deng, Zhi Ding

Signal processing over single-layer graphs has become a mainstream tool owing to its power in revealing obscure underlying structures within data signals. However, many real-life data sets and systems are characterized by more complex interactions among distinct entities, which may represent multilevel interactions that are harder to be captured with a single-layer graph, and can be better characterized by multilayer graph connections. Such multilayer or multilevel data structures can be more naturally modeled by high-dimensional multilayer graphs (MLGs). To generalize traditional graph signal processing (GSP) over MLGs for analyzing multilevel signal features and their interactions, this presentation will introduce a tensor-based framework of MLG signal processing (M-GSP). Specifically, we will first introduce core concepts of M-GSP and study the properties of MLG spectral space, followed by fundamentals of MLG-based filter design. To illustrate novel aspects of M-GSP, we will further provide the example applications to demonstrate the efficacy and benefits of applying MLGs and M-GSP in practical scenarios. Finally, we will discuss its correlations to graph neural networks and its potential for tensor-based MLG neural networks (MGNN).

3 - Combining Knowledge Graph and Large Language Model: use Cases and Challenges

Lili Zhang, Hewlett Packard Enterprise, Alpharetta, GA, United States

Both Knowledge Graph and Large Language Model have generated significant values across industries and daily life. Knowledge Graph is efficient in identifying the patterns from the relationships. Large Language Model mimics human language understanding and response behaviors. In this session, we explore the use cases (e.g., webpage phishing detection) and challenges of combining Knowledge Graph and Large Language Model.

SD12

Summit - 332

Digital Innovation and Social Media

Flash Session

Flash

Chair: Somak Paul, California State University, Alameda, CA, United States

1 - Effects of digital orientation on innovation performance: a mediating role of digital capabilities

Xian Zhang, School of Management, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Wenyi Luo, Feifei Guo, Peng Wang

In the era of the digital economy, digital orientation is considered a crucial development strategy for firms. Despite the predominant focus on the effect of digital orientation on innovation performance, there is a lack of research exploring the mechanisms through which digital orientation enables innovation performance. Drawing on dynamic capability theory as the core theoretical foundation, this study investigates the Interrelationships of digital orientation, digital capabilities, and innovation performance from a dynamic evolutionary perspective and examines the mediating effect of digital capabilities and the moderating effect of environmental uncertainty. We collected survey data from 368 Chinese firms and used hierarchical linear regression to test our hypotheses. The empirical results show that digital orientation and the three sub-dimensions of digital capabilities (digital sensing capabilities, digital integrative capabilities, and digital network capabilities) all promote innovation performance; digital capabilities have a positive mediating effect on the relationship between digital orientation and innovation performance; environmental uncertainty moderates the relationship between digital capabilities and innovation performance, but does not moderate the relationship between digital orientation and innovation performance or between digital orientation and digital capabilities. This study reveals the micro-mechanism between digital orientation and innovation performance and constructs a theoretical framework of "digital orientation-digital capabilities-innovation performance", which is a beneficial extension of dynamic capability theory in the context of digitalization. Furthermore, the research findings encourage firms to enhance their digital capabilities by adopting digitally-oriented strategies to cope with changing environments and improve their innovation performance.

2 - Understanding the Role of Developers' Social Interaction within Open Source Software Platforms on Strategic Directions for Digital Innovation

Sidi Zhao, Auburn University, Auburn, AL, United States, Pei Xu, Uzma Raja

Firms proactively invest in digital innovation to strategically allocate IT resources and boost business performance. Upper Echelons Theory suggests that top managers' characteristics influence strategic directions, but many lack the technical expertise for effective decision-making. Digital innovations require new organizational forms to manage complex relationships and interactions among actors at various knowledge and learning levels. Many firms leverage open innovation platforms like open-source software (OSS) communities to access external knowledge and resources. While the benefits of open innovation are well-documented, the impact of social interactions among OSS developers on firms' digital innovation efforts is underexplored. We posit that social dynamics within OSS communities significantly shape knowledge diffusion across firms. Utilizing a heterogeneous diffusion model, we develop a framework to understand how social context impacts innovation adoption through social proximity, infectiousness, and susceptibility factors. We construct the social network among firms' developers on GitHub and the patent network of artificial intelligence (AI) and Blockchain-related patents from the US Patent and Trademark Office (USPTO). Blockchain and AI, as typical forms of digital innovation, have substantial potential to impact business activities. Based on link prediction and panel fixed effect estimation on longitudinal datasets from 2010 to 2023, our empirical investigation contributes to social network analysis by addressing the complexities of predicting firm network ties through individual social interactions.

The findings offer actionable insights into how developers influence firm strategy from the bottom up, helping firms maintain strategic flexibility and stay attuned to innovation trends among market peers.

3 - Analyzing the effects of knowledge management on organizational performance through knowledge utilization and sustainability

Faezeh Mohaghegh, Alfred University, Alfred, NY, United States, Halil Zaim, Vladimir Dzenopoljac, Aleksandra Dzenopoljac, Nick Bontis

In today's turbulent business environment, gaining sustainable competitive advantage requires organizations to manage knowledge capabilities effectively and use resources efficiently. The primary objective of this study is to explore the relationship among knowledge management (KM) processes, including knowledge generation, knowledge codification, and knowledge sharing, with knowledge utilization, sustainability, and organizational performance. The research mainly examines both the direct effect of KM on organizational performance and its indirect effect through sustainability and knowledge utilization. A theoretical framework is introduced and tested using data gathered from companies within the services sector in Kuwait. The results indicate that knowledge utilization and sustainability fully mediate the relationship between KM and organizational performance. Both knowledge utilization and sustainability exert direct and positive effects on organizational performance while being directly influenced by KM. Accordingly, it is suggested that the firms enhance their performance by effectively managing knowledge capabilities appropriately, utilizing knowledge, and investing in sustainability and organizational resources.

4 - Decentralized Decision-Making Framework for Managing Product Rollovers in the Semiconductor Industry Under Uncertainty.

Carlos Leca, North Carolina State University, Raleigh, NC, United States, Karl Kempf, Reha Uzsoy

In competitive industries, firms seek to gain competitive advantage through new product introductions and the retirement of existing products, a process termed *'product rollover.'* Navigating an effective rollover strategy is a complex task, that can jeopardize a firm's financial stability if not managed adeptly. Firms face the challenge of *resource allocation* across diverse divisions, each serving distinct market niches, possessing proprietary information, and *competing* for the firm's resources.

In prior research, we developed a *decentralized resource allocation mechanism* that coordinates agents involved in product rollover strategy execution within a reasonable computational time. However, this mechanism was only tested under deterministic conditions. To address this limitation, our current study integrates the decentralized model into a rolling horizon framework capable of accommodating *parameter uncertainty*. Furthermore, we propose an adapted framework fostering *strategic conduct* among agents, allowing them to hedge against uncertainty by signaling oversized estimated allocation requirements. This perturbation from the estimated requirements leads to increased resource scarcity impacting the resource cost in future periods and hence affecting the performance of other agents.

Finally, we applied a *reinforcement learning* technique to determine the agent's optimal buffering level given the current state of the rollover campaign. The resulting policy matrices shed light on situations where an agent's optimal decision, driven by self-interested objectives, may not align with the firm's overall objectives.

5 - Adapting to Adversity: The Impact of Data Breaches on the Disclosure of Digital Strategic Initiatives

HILAL PATACI, University of Texas at San Antonio, San Antonio, TX, United States, T Ravichandran, Nishtha Langer

Digital transformation is fundamentally reshaping how organizations operate and deliver value to their stakeholders. Central to this transformation are digital strategic initiatives, which encompass specific projects and programs undertaken to leverage digital technologies and achieve strategic objectives. These initiatives often entail implementing advanced security measures to safeguard sensitive data and prevent unauthorized access. Understanding how these initiatives are influenced by factors such as data breaches is crucial for organizations navigating the complex challenges of digital transformation and cybersecurity. Drawing on disclosure and institutional theories, we explore mechanisms driving organizations' voluntary disclosure decisions post-data breaches, considering their access to privileged information about breaches in their competitors and upstream suppliers. Institutional theory suggests that organizations face mimetic pressures, compelling them to emulate successful, non-breached peers. Disclosure theory emphasizes meeting shareholders' information needs, prompting organizations to disclose more nonproprietary information under mandatory disclosure requirements. Organizations acting as customers of breached upstream suppliers often access privileged information on breach specifics, prompting them to refine their disclosure strategies to assuage stakeholder concerns and minimize uncertainty by disclosing less proprietary information. Conversely, organizations associated with breached competitors, with limited access to privileged information, face heightened pressure to conform to industry standards. They enhance their disclosure efforts to maintain shareholder trust, even without precise breach information, by disclosing more proprietary information. Our findings highlight that access to privileged information leads to more conservative voluntary disclosure behavior. Further, organizations with prior breach experience possess deeper insights into breach impacts, shaping their disclosure strategies accordingly.

6 - Partial Persuasive Marketing on Service Operations Platforms

Yang XIA, City University of Hong Kong, Hong Kong, China, People's Republic of, Pengfei Guo, Qiaochu He, Yulan Wang

This study investigates the strategic use of partial persuasive marketing on an online service platform under uncertain service quality. We model the partial persuasive marketing with “targeting technology”, assuming that the platform strategically communicates service quality to a subset of customers from the perspective of Bayesian persuasion. A queuing game model is developed to determine the platform’s optimal information strategy, aiming to tune the online service recommendation algorithm’s accuracy. The algorithm commits to binary signal strategies based on service quality, influencing informed customers’ beliefs. All customers then decide whether to join the platform. Interestingly, at a moderate market size, we discover that adjusting the beliefs of partial customers with targeting technology can achieve the same effective arrival rate as adjusting the beliefs of all customers. We show that the platform can maximize throughput by fine-tuning positive signal accuracy or to enhance social welfare by obfuscating negative signals. We demonstrate that a unique targeting strategy that maximizes platform throughput can be identified, and partial persuasive marketing is better in a large market size since it is more difficult to adjust the beliefs of all customers than to adjust the beliefs of a subset of customers. To maximize social welfare, we demonstrate that the optimal targeting strategy and information disclosure level can significantly enhance collective payoffs, particularly when market sizes are unconstrained. This insight challenges the intuition that strategically communicating with all potential customers is invariably the preferred strategy, highlighting the benefits of well-designed and targeted information campaigns.

7 - Operational Inefficiencies and Misinformation

Jingyi Sun, Stevens Institute of Technology, Hoboken, NJ, United States, Feng Mai, Muer Yang

This study examines the relationship between operational factors such as waiting time in queues and online misinformation spread using a unique dataset combining granular measures from geospatial data, social media content, and demographic information. Employing mixed research methods, the findings suggest that operational variables are associated with increased sharing of misinformation, with significant heterogeneity based on demographics. The research introduces a novel operational factor into the study of information dynamics on social media.

8 - Fairness as a Heuristic in Socially Responsible Decision-Making: An Experimental Study

Somak Paul, California State University, Hayward, CA, United States

The fairness heuristic is an effective lever as it allows people to free up their cognitive capacity in situations that could otherwise overwhelm their cognitive and attentive resources. The perception of fair treatment leads people to respond cooperatively to the demands and requests of others. On the other hand, if they perceive to have been treated unfairly, individuals reciprocate by rejecting cooperation and embrace self-interested orientation. In this research we study how the decision-maker in a supplier firm faces a complicated situation where they not only need to focus on their profit motif but also need to make investment decisions for workplace safety compliance. Through a controlled experiment we observe that the perception of procedural and distributional fairness in a buyer's ordering characteristics impact how the supplier firm wants to orient itself in fulfilling its social responsibility. Additionally, we note that the financial reliability of the buyer in upholding the contract also impacts the supplier's investment in building a safe workplace. Our research illuminates that buyers should use their fairness lever carefully to help establish socially responsible operations in their supplier's workplace.

SD13

Summit - 333

Healthcare Resource Management

Flash Session

Contributed

Chair: Jens Brunner, Technical University of Denmark, Akademivej, Kgs. Lyngby, 2800, Denmark

1 - Optimizing Ambulance Allocation to On-Campus and Free-Standing Emergency Departments Using a Multistage Stochastic Programming Approach

Jorge Acuña, Universidad Adolfo Ibáñez, Viña del Mar, Chile, Daniela Cantarino, Jose Zayas-Castro, Michael Lozano Jr.

Emergency department (ED) overcrowding is a significant issue in the US healthcare system, leading to adverse patient outcomes. ED visits have surged by over 60% since 1997, surpassing 130 million annually, prompting the construction of free-standing EDs across the country. However, patient transfers between on-campus and free-standing EDs exacerbate issues like ambulance diversion and prolonged EMS wait times. The literature has shown that overcrowding stems from systemic dysfunction rather than individual inefficiencies. We propose a stochastic mixed-integer programming model for allocating ambulances to on-campus and free-standing EDs over a finite planning horizon to address this. Our model considers patient conditions, priorities, real-time routing, and ED capacities to minimize travel and wait times. We developed an enhanced progressive hedging algorithm (EPHA) to analyze a real-size scenario in a county in Florida. While commercial solvers may find optimal solutions faster for small instances, they struggle with larger scenarios. In contrast, the EPHA delivers near-optimal solutions within 15 minutes for real scenarios. The preliminary results show that ignoring current and future utilization levels can turn free-standing EDs into bottlenecks. Therefore, a centralized decision system for allocation policies can reduce average patient wait times by up to 45% (27 minutes), improve EMS and ED efficiency, and provide robust alternatives during demand surges. Proper utilization of the combined capacities of free-standing and on-campus EDs ensures prompt patient care, even during high demand. Sensitivity analysis on primal and dual multipliers highlights the significance of the penalty update method in enhancing EPHA's accuracy and computation time.

2 - Aggregate surgery capacity planning with consideration of self-cancellation

Xue Yang, Simon Fraser University, Burnaby, BC, Canada, Zhe Zhang, Yang Wang, Abraham Punnen

Patient self-cancellation in day surgery centers (DSCs), as common as the cancellation because of insufficient surgery resources or emergency patients, leads to waste of surgery resources and increase of the hospital costs. However, current healthcare capacity planning methods usually only consider the latter. To address such a shortcoming, we examine a long-term sequential surgery capacity planning problem in which the DSC allocates operating room hours and surgical intensive unit care (SICU) bed days to different specialties with consideration of the time-dependent self-cancellation-related waiting risk. We formulate this problem as a Markov decision process, explicitly taking into account patient length-of-stay in SICU beds, self-cancellation chances and elective patient census. We solve the model by approximate dynamic programming: variable aggregation-based Lagrangian relaxation and constraint-generation-based heuristics. We also propose a heuristic policy in which patients in different specialties are viewed as single-type patients. To illustrate the efficiency of our approaches and draw managerial insights, we apply our solutions to a two-year data set from the day surgery center in a large public hospital in Sichuan, China. We examine the quality of our approaches and compare their solutions with the proposed heuristic policy. We also show good performances of our approaches in the complex environment through the extensive numerical analysis.

3 - Operating Room Scheduling under Human Resource

Xiajie Yi, IÉSEG School of Management, Lille, France, Christine DI MARTINELLY

Effective Operating Room (OR) scheduling holds significant financial implications for hospitals, particularly when dealing with inpatient procedures at the operational level. However, the duration of a surgery is often uncertain and it is rather challenging to accurately predict. Estimates can vary significantly across different departments, leading to various extents of underestimations. As a result, it exacerbates nurse shortage issue in hospitals. To address this, we propose an approach to build robust surgery schedules in an OR theatre made of several rooms with different specialties and sharing nurses. Our method utilizes a two-stage stochastic optimization model. In the first stage, we determine the number of nurses working each day, while the second stage focuses on creating the detailed schedule. Additionally, we adopt the sampling average approximation (SAA) technique to address uncertainties in surgery duration. Based on the samples generated from real-life data, the simulation results imply that i) the surgery schedule could be fully scheduled with fewer number of nurses; ii) the nurse shortage/absence issue can be potentially avoided as we can accommodate uncertain duration better.

4 - Planning surgeries and staff simultaneously subject to scheduling flexibility

Jens Brunner, Technical University of Denmark, Kgs. Lyngby, Denmark

Physicians and operating rooms are two of hospitals' most important, expensive, and scarce resources. Therefore, effective and efficient scheduling of these resources is among hospitals' most relevant planning tasks. The decisions to be made on the daily planning level regarding sequencing patients’ surgeries and assigning appropriate staff to surgeries. Since there are several interdependencies between surgery schedules and physician rosters, it is a meaningful approach to consider both planning problems within one integrated optimization problem. We provide mixed-integer programming models within a column generation algorithm that solves both scheduling problems. We create schedules for surgeries and operating room staff. We use test data based on a real-world dataset to provide meaningful insights. The algorithm leads to optimal solutions in more than 75% of all test cases and solves the problem efficiently within a desirable amount of time. We further evaluate our algorithm with respect to different aspects of flexibility in the context of surgery and staff scheduling and generate key insights about their interdependencies.

SD14

Summit - 334

Climate Financial Risk

Invited Session

Finance

Chair: Kay Giesecke, Stanford University, Stanford, CA, United States

1 - How Much Sustainability is Really in Stock Prices?

Enrica Archetti, Stanford University, Stanford, CA, United States, Eva Lütkebohmert, Markus Pelger, Marcus Rockel

We show that financial markets do not price environmental risk once we control for conventional sources of risk. In a comprehensive empirical study we use machine learning to predict stock returns with a large set of conventional firm characteristics and climate characteristics. Our battery of machine learning methods allows us to capture complex dependencies and control for a large number of variables. When using only climate variables, they predict stock returns. However, this predictability disappears when controlling for conventional sources of risk. These results are robust to the choice of machine learning method. Thus, sustainability characteristics do not represent a new source of financial risk but merely repackage existing risk factors.

2 - Physical Climate Risk and Small Business Defaults

Alec Madayan, Stanford University, Stanford, CA, United States

This study examines the causal impact of physical climate risk on small business loan defaults in the United States, a critical factor for insurance pricing. Initially focusing on Hurricane Harvey, we analyze its effect on small business defaults while controlling for various confounding variables. Our findings demonstrate a significant causal relationship between climate risk and loan defaults. Notably, the timing of defaults is influenced by specific loan characteristics and the severity of the climate impact on the region of the issuer. We employ both linear and nonlinear modeling approaches, along with clustering algorithms, to estimate these effects and conduct robustness checks to ensure the reliability of our results. Our methodology is then systematically applied across various climate events, offering a comprehensive understanding of the interplay between climate risk and small business financial stability. This research provides valuable insights for insurers and policymakers, aiding in the development of more accurate risk assessment models and pricing strategies.

3 - Physical Climate Risk for Mortgage Markets

Johannes Fuest, Stanford University, Stanford, CA, United States

We develop causal machine learning estimators for point processes tracking the timing of correlated events. We apply these estimators to evaluate the impact of physical climate risk on the behavior of individual mortgage borrowers and agency mortgage-backed securities (MBS). We find that natural disasters such as hurricanes have significant causal effects on mortgage borrower behavior including voluntary and involuntary prepayments. The effects are found to vary greatly with borrower characteristics including leverage, geography, and credit score, as well as macro-economic conditions including interest rates. However, the geographic diversity of agency MBS collateral loan pools greatly mitigates these causal effects, basically shielding MBS investors from physical climate risk. These findings have important implications for the disclosure of climate-related risks by financial institutions and the optimal structuring of MBS pools by agencies such as Fannie Mae.

4 - Transportation Marketplace Rate Forecast Using Signature Transform

Xinyu Li, UC Berkeley, Berkeley, CA, United States

Freight transportation marketplace rates are typically challenging to forecast accurately. In this work, we have developed a novel statistical technique based on signature transforms and have built a predictive and adaptive model to forecast these marketplace rates. Our technique is based on two key elements of the signature transform: one being its universal nonlinearity property, which linearizes the feature space and hence translates the forecasting problem into linear regression, and the other being the signature kernel, which allows for comparing computationally efficiently similarities between time series data. Combined, it allows for efficient feature generation and precise identification of seasonality and regime switching in the forecasting process. An algorithm based on our technique has been deployed by Amazon trucking operations, with far superior forecast accuracy and better interpretability versus commercially available industry models, even during the COVID-19 pandemic and the Ukraine conflict. Furthermore, our technique is able to capture the influence of business cycles and the heterogeneity of the marketplace, improving prediction accuracy by more than fivefold, with an estimated annualized saving of $50 million.

SD15

Summit - 335

Innovations in Data-driven Marketplaces

Invited Session

Revenue Management and Pricing

Chair: Yiding Feng, University of Chicago Booth School of Business, Chicago, IL, N/A, United States

1 - Dynamic Pricing of Split Stays

Yi Zhuang, Baruch College, The City University of New York, New York City, NY, United States, Arash Asadpour, Yuan-Mao Kao

Traditionally, hotel bookings are listed based on the search for check-in and check-out dates for a room. Recently, Airbnb introduced an option to book split stays for two listings. However, this option can incur inconvenience when a traveler needs to relocate to a new lodging during their stay. Thus, pricing considerations for split stays are different from those for single reservations, and relocating discounts might be in order. We develop a dynamic pricing framework that provides the optimal pricing policy for split stays as an alternative to single reservations. We study how the discounts for split stays change over time and in different availability scenarios. We find that split stays play a role in balancing the inventory of the rooms when it is about to deviate. We also study the revenue gains resulting from offering these new products.

2 - Regulation of Algorithmic Collusion

Chenhao Zhang, Northwestern University, Evanston, IL, United States, Jason Hartline, Sheng Long

Consider sellers in a competitive market that use algorithms to adapt their prices from data that they collect. In such a context it is plausible that algorithms could arrive at prices that are higher than the competitive prices and this may benefit sellers at the expense of consumers (i.e., the buyers in the market). This paper gives a definition of plausible algorithmic non-collusion for pricing algorithms. The definition allows a regulator to empirically audit algorithms by applying a statistical test to the data that they collect. Algorithms that are good, i.e., approximately optimize prices to market conditions, can be augmented to collect the data sufficient to pass the audit. Algorithms that have colluded on, e.g., supra-competitive prices cannot pass the audit. The definition allows sellers to possess useful side information that may be correlated with supply and demand and could affect the prices used by good algorithms. The paper provides an analysis of the statistical complexity of such an audit, i.e., how much data is sufficient for the test of non-collusion to be accurate.

3 - Bayesian Persuasion with a Learning Agent

Tao Lin, Harvard University, Cambridge, MA, United States, Yiling Chen

We study a repeated Bayesian persuasion problem (and more generally, any generalized principal-agent problem with complete information) where the principal does not have commitment power and the agent uses algorithms to learn to respond to the principal's signals. We reduce this problem to a one-shot generalized principal-agent problem with an approximately-best-responding agent. This reduction allows us to show that: if the agent uses contextual no-regret learning algorithms, then the principal can guarantee a utility that is arbitrarily close to the principal's optimal utility in the classic non-learning model with commitment; if the agent uses contextual no-swap-regret learning algorithms, then the principal cannot obtain any utility significantly more than the optimal utility in the non-learning model with commitment. The difference between the principal's obtainable utility in the learning model and the non-learning model is bounded by the agent's regret (swap-regret). If the agent uses mean-based learning algorithms (which can be no-regret but not no-swap-regret), then the principal can do significantly better than the non-learning model. These conclusions hold not only for Bayesian persuasion, but also for any generalized principal-agent problem with complete information, including Stackelberg games and contract design.

4 - Are Bounded Contracts Learnable and Approximately Optimal?

YURONG CHEN, Peking University, Beijing, China, People's Republic of, Zhaohua Chen, Xiaotie Deng, Zhiyi Huang

This paper considers the hidden-action model of the principal-agent problem, in which a principal incentivizes an agent to work on a project using a contract. We investigate whether contracts with bounded payments are learnable and approximately optimal. Our main results are two learning algorithms that can find a nearly optimal bounded contract using a polynomial number of queries, under two standard assumptions in the literature: a costlier action for the agent leads to a better outcome distribution for the principal, and the agent's cost/effort has diminishing returns. Our polynomial query complexity upper bound shows that standard assumptions are sufficient for achieving an exponential improvement upon the known lower bound for general instances. Unlike the existing algorithms which relied on discretizing the contract space, our algorithms directly learn the underlying outcome distributions. As for the approximate optimality of bounded contracts, we find that they could be far from optimal in terms of multiplicative or additive approximation, but satisfy a notion of mixed approximation.

5 - Screening Signal-Manipulating Agents via Contests

Xiaoyun Qiu, Northwestern University, Evanston, IL, United States, Yingkai Li

We study the design of screening mechanisms subject to competition and manipulation. A social planner has limited resources to allocate to multiple agents using *only* signals manipulable through unproductive effort. We show that the welfare-maximizing mechanism takes the form of a contest and characterize the optimal contest. We apply our results to two settings: either the planner has one item or a number of items proportional to the number of agents. We show that in both settings, with sufficiently many agents, a winner-takes-all contest is never optimal. In particular, the planner always benefits from randomizing the allocation to some agents.

SD16

Summit - 336

Bounds, Heuristics, and Personalized Assortments

Invited Session

Revenue Management and Pricing

Chair: Guillermo Gallego, Chinese University of Hong Kong Shenzhen, Shenzhen, N/A

Co-Chair: Pin Gao, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

1 - Comparative Analysis of Parsimonious Discrete Choice Models for Prediction and Assortment Optimization

Guillermo Gallego, Chinese University of Hong Kong Shenzhen, Shenzhen, China, People's Republic of, Anran Li, Pengyu Zuo

The bias-variance tradeoff suggest the use of parsimonious discrete choice models for small to moderately-sized datasets. This study assesses five such models, focusing on predictive accuracy and assortment optimization, both with and without cardinality constraints. We compare the established Exponomial, BAM, GAM, and RCS models against a new variant, the G-RCS model – a generalization of RCS. The G-RCS model extends its functionality to include shadow attractions for unavailable products and offers a closed-form representation based on first-and-last choice probabilities. The talk will elaborate on the G-RCS model's attributes and conduct a numerical comparison of all five models, utilizing both synthetic and real-world industrial datasets.

2 - LP-Based Control for Network Revenue Management under Markovian Demands

Haixiang Lan, Columbia University, New York, NY, United States

We study the network revenue management (NRM) problem, where the seller irrevocably accepts or rejects customers' requests to maximize the total revenue over a finite time horizon subject to limited resources. We consider the setting where customer arrivals are dependent and follow the Markov chain choice model. We propose a deterministic linear program (DLP), a fluid approximation of the original problem, and provide an upper bound for the expected revenue of the optimal policy. We design a static probabilistic allocation (SPA) policy which accepts the request for each product with a certain probability and show that the randomized policy is asymptotically optimal. We show that the regret of the randomized policy, defined as the difference between our policy's expected revenue and the DLP's optimal value, scales in the square root of the time horizon.

When the parameters of our model, the initial arrival probabilities of each product and the transition probabilities between products, are unknown \textit{a priori}, the seller needs to simultaneously (i) learn customers' preferences based on the collected purchase and transition data, and (ii) maximize the cumulative revenue by making online acceptance or rejection decisions while satisfying the resource constraints. To achieve such an exploration-exploitation trade-off, we design an online probabilistic allocation (OPA) policy that achieves a near-optimal order of regret up to a logarithmic factor under some mild conditions. We also conduct numerical experiments to validate the performance of the proposed policies.

3 - Simple is Enough: A Cascade Approximation for Attention-Based Satisficing Choice Models

Chenhao Wang, The Chinese University of Hong Kong, ShenZhen, China, People's Republic of, Pin Gao, Yicheng Liu, Zizhuo Wang

Empirical evidence suggests that consumers commonly focus their attention on a subset of available products and evaluate them in batches to identify a satisfactory option. To capture this phenomenon, we introduce the attention-based satisficing choice rule, which encompasses special cases such as the sequential MNL (e.g., Gao et al. 2021), click-based MNL (e.g., Aouad et al. 2019), and random consideration set models (e.g., Gallego and Li 2017). Through an empirical investigation employing data sourced from Expedia, we provide evidence that special cases of the proposed model exhibit a notable advantage in terms of predictive accuracy when compared to the mixed MNL. Notwithstanding the NP-hardness of finding the revenuemaximizing assortment and estimating certain parameters for the proposed model, we demonstrate that it can be approximated by a simple cascade model (e.g., Kempe and Mahdian 2008) with substantially fewer parameters. Specifically, we establish that the overall likelihood of purchasing from any given assortment under the proposed model can be estimated within a certain range, multiplied by that in the cascade model; moreover, by utilizing the readily computable optimal assortment derived from the approximated model as a heuristic, we show that the worst-case revenue given partial information is at least 3/8 of that obtained from an optimized assortment under the best parameter configuration. Finally, based on the technique established in this study, we extend the analysis to the constrained assortment optimization problem, the joint assortment and pricing problem, and the categorized attention-based assortment optimization problem.

SD17

Summit - 337

AI and Operations

Invited Session

Revenue Management and Pricing

Chair: Hannah Li, Columbia, New York, NY, United States

Co-Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

1 - Can ChatGPT do it? Identifying product complements using pre-trained large language models.

Srikant Jagabathula, NYU, New York, NY, United States

With large pre-trained language models (LLMs), we appear to be closer than ever to achieving universal prediction engines. For every prediction problem we aim to solve, it is natural to wonder if the commercially available LLMs can already handle it. Many practitioners are leveraging these models' predictions for critical business applications. But is it appropriate to rely on these models for such essential tasks? Beyond hallucinations, what other concerns should we consider? In this work, we systematically address these and various other questions.

2 - Causal Inference for Human-Language Model Collaboration

Paramveer Dhillon, University of Michigan, Ann Arbor, MI, United States

We examine the collaborative dynamics between humans and language models (LMs), where the interactions typically involve LMs proposing text segments and humans editing or responding to these proposals. Productive engagement with LMs in such scenarios necessitates that humans discern effective text-based interaction strategies, such as editing and response styles, from historical human-LM interactions. This objective is inherently causal, driven by the counterfactual `what-if' question: how would the outcome of collaboration change if humans employed a different text editing/refinement strategy? A key challenge in answering this causal inference question is formulating an appropriate causal estimand: the conventional average treatment effect (ATE) estimand is inapplicable to text-based treatments due to their high dimensionality. To address this concern, we introduce a new causal estimand-- Incremental Stylistic Effect (ISE), which characterizes the average impact of infinitesimally shifting a text towards a specific style, such as increasing formality. We establish the conditions for the non-parametric identification of ISE. Building on this, we develop CausalCollab, an algorithm designed to estimate the ISE of various interaction strategies in dynamic human-LM collaborations. Our empirical investigations across three distinct human-LM collaboration scenarios reveal that CausalCollab effectively reduces confounding and significantly improves counterfactual estimation over a set of competitive baselines.

3 - Optimal Experimentation for Learning Personalized Policies Across Locations

Spyros Zoumpoulis, INSEAD, Fontainebleau, France, Georgina Hall, Stefanos Poulidis

Firms wish to learn personalized policies for customers in heterogeneous yet related locations to maximize their monetary gains. To do this, they conduct experiments at each location to estimate the parameters of a customer response function. A crucial decision is which action to assign to each participant in the experiment, especially when a participant can only be assigned one action or there are budget constraints. The existing experimentation methodology considers locations and experiments individually. In this work, we leverage the relationship between locations in the experimentation problem to learn more profitable policies by proposing novel estimators and a semidefinite programming approach.

4 - Planning Adaptive Experiments: A Math Programming Approach

Ethan Che, Columbia Business School, New York, NY, United States, Daniel Jiang, Hongseok Namkoong, Jimmy Wang

Adaptive experimentation can improve statistical power significantly, but typical algorithms overlook important issues that arise in practice: multiple objectives, non-stationarity, batched/delayed feedback, constraints, and personalization. Moving away from developing bespoke algorithms for each setting, we present a mathematical programming view of adaptive experimentation that can flexibly incorporate a wide range of objectives, constraints, and statistical procedures. By formulating a dynamic program in the batched limit, our modeling framework enables the use of scalable optimization methods (e.g., SGD and auto-differentiation) to solve for treatment allocations. To spur algorithmic progress, we build a suite of benchmark problems based on hundreds of real A/B tests at ASOS that model key practical issues such as non-stationarity, personalization, multi-objectives, and constraints. Our empirical results show standard Thompson sampling-based policies fail to reliably improve upon static designs, and demonstrate the effectiveness of a simple planning approach.

SD18

Summit - 338

Learning and Causal Inference

Invited Session

Revenue Management and Pricing

Chair: Dennis Zhang, Washington University in St Louis, ST LOUIS, MO, United States

Co-Chair: Heng Zhang, Arizona State University, Tempe, AZ, United States

1 - Deep Learning for Policy Targeting with Continuous Treatment

Zhiqi Zhang, Washington University in St. Louis, St. Louis, MO, United States, Zhiyu Zeng, Ruohan Zhan, Dennis Zhang

Experimentation has become a cornerstone technique in product development and operations management for technology companies. While many experiments focus on binary treatments (hence the term A/B testing), these companies increasingly conduct experiments to determine the optimal level of continuous treatment variables. Platforms often approach this problem by discretizing the continuous variable, testing each level as one treatment, and selecting the best level. While discretizing a continuous treatment into different levels in an experiment is simple to implement, this approach has several shortcomings. First, it cannot provide any statistical guarantees on the treatment levels beyond the tested ones. Second, since the experiment cannot provide any statistical guarantees on the untested treatments, it cannot provide any performance guarantees on the derived policies. Third, this approach ignores the important dependencies of treatment effects on often high-dimensional and influential user features. We propose a double-machine-learning framework that leverages personal characteristics to determine the optimal treatment level for each customer. Unlike traditional discrete experiments, our approach provides theoretical guarantees on revenue and enables the identification of the optimal treatment level without relying on discrete levels. We rigorously validate our approach using a random field experiment on a large-scale video-sharing platform and conduct synthetic studies for further validation.

2 - Adaptive Experimentation when You Can&Rsquo;T Experiment: Efficient Encouragement Designs via Experimental Design

Lalit Jain, Foster School of Business, University of Washington, Seattle, WA, United States, Yao Zhao, Kwang-Sung Jun, Tanner Fiez

Often online services cannot directly assign users to specific control or treatment experiences either for business or practical reasons. In these settings, naively comparing treatment and control groups that may result from self-selection can lead to biased estimates of underlying treatment effects. Instead, online services can employ a properly randomized encouragement that incentivizes users toward a specific treatment. Our methodology provides online services with an adaptive experimental design approach for learning the best-performing treatment for such encouragement designs. We consider a more general underlying model captured by a linear structural equation and formulate pure exploration linear bandits in this setting. Though pure exploration has been extensively studied in standard adaptive experimental design settings, we believe this is the first work considering a setting where noise is confounded. Elimination-style algorithms using experimental design methods in combination with a novel finite-time confidence interval on an instrumental variable style estimator are presented with sample complexity upper bounds nearly matching a minimax lower bound. Finally, experiments are conducted that demonstrate the efficacy of our approach.

3 - Harnessing Large Language Models for Market Research: a De-Biasing Approach

Mengxin Wang, The University of Texas at Dallas, Richardson, TX, United States, Dennis Zhang, Heng Zhang

This study investigates the potential of large language models (LLMs) as data augmenters in market research, focusing on their ability to generate conjoint choice data traditionally collected through expensive and resource-intensive surveys. Conjoint analysis, a staple in market research, depends on surveys that elicit consumer preferences for products characterized by various attributes, facilitating the estimation of preference models and utility parameters critical for product development. Our research examines whether LLMs can replicate the intricate decision-making processes inherent in these analyses, offering a cost-effective alternative to human subject surveys.

Past research primarily focused on the empirical efficacy of LLM-generated data in enhancing data quality and alignment with human responses through prompt engineering. Our work diverges by developing a novel statistical de-biasing methodology, which aligns any LLM-generated data with human responses. This method leverages minimal human responses to detect correlations between LLM responses and human choices in specific contexts. The de-biasing performance strengthens when employing advanced prompt engineering techniques such as the Chain-of-Thought method.

Theoretical evaluations of our approach demonstrate significant statistical advantages over traditional methods, promising to reduce logistical barriers and democratize access to high-quality market research tools. These findings not only enhance the understanding of LLM capabilities and limitations but also pave the way for broader applications in market research, potentially revolutionizing the field.

4 - Deep Causal Inequalities: Demand Estimation in the Differentiated Products Markets

Jiding Zhang, Arizona State University, Tempe, AZ, United States, Edvard Bakhitov, Amandeep Singh

In this paper, we borrow from the literature on partial identification and propose the Deep Causal Inequalities (DeepCI) estimator that overcomes both these issues. Instead of relying on observed labels, we use inferred moment inequalities from the observed behavior of agents in the data, which allows us to factor out the endogenous component and nonparametrically model the underlying utility. We demonstrate that DeepCI is consistent under mild restrictions and derive the estimation bounds in the projected root mean squared error norm. The approach is flexible and can be applied to a large variety of hypothesis spaces.

SD19

Summit - 339

Exploration, Experimental Design, and Algorithmic Decision Making

Invited Session

Revenue Management and Pricing

Chair: Jinglong Zhao, Boston University, Boston, MA, United States

Co-Chair: Yunzong Xu, University of Illinois Urbana-Champaign, Champaign, IL, United States

1 - Harnessing the Continuous Structure: Improved Algorithms in Online Contract Design

Shiliang Zuo, UIUC, Savoy, IL, United States

This work studies online learning algorithms for repeated principal-agent problems, where the principal seeks to learn the optimal contract that maximizes utility through repeated interactions, without prior knowledge of the agent's type (i.e., cost and production functions). By leveraging economic structures, specifically the first-order approach, we show that learning the optimal contract can be reduced to a Lipschitz bandit problem. In the special case of learning an optimal linear contract, this reduction simplifies further to a 1-dimensional continuum-armed bandit problem, sharing the same structure as the dynamic pricing problem.

2 - A/B tests: non-stationarity and beyond

Yuhang Wu, University of California, Berkeley, Berkeley, CA, United States

A/B tests have been used at scale by data-driven enterprises to guide decisions. Meanwhile, non-stationarities can often arise in key business metrics. In this work, we discuss challenges of non-stationary A/B tests and discuss how to address them. We also discuss some extensions of other aspects in A/B tests with additional constraints or goals, This includes joint work with Zeyu Zheng and Amazon.

3 - Statistical Properties of Robust Satisficing

Zhiyi Li, Peking University, Beijing, China, People's Republic of, Yunbei Xu, Ruohan Zhan

The Robust Satisficing (RS) model is an emerging approach to robust optimization, offering streamlined procedures and robust generalization across various applications. However, the statistical theory of RS remains unexplored in the literature. Our work fills in the gap by comprehensively analyzing the theoretical properties of the RS model. Notably, the RS structure offers a more straightforward path to deriving statistical guarantees compared to the seminal Distributionally Robust Optimization (DRO), resulting in a richer set of results. In particular, we establish two-sided confidence intervals for the optimal loss without the need to solve the optimization problem explicitly. We further provide finite-sample generalization error bounds for the RS optimizer. Importantly, our results extend to scenarios involving distribution shifts, where discrepancies exist between the sampling and target distributions. Our numerical experiments show that the RS model consistently outperforms the baseline empirical risk minimization in small-sample regimes and under distribution shifts. Furthermore, compared to the DRO model, the RS model exhibits lower sensitivity to hyperparameter tuning, highlighting its practicability for robustness considerations.

4 - Can large language models explore in-context?

Akshay Krishnamurthy, Microsoft Research, New York, NY, United States

An emergent paradigm in machine learning is in-context learning, where one uses a pre-trained foundation model to solve a problem by specifying the problem description and data entirely within the model's context, with no updates to the model parameters. We study in-context reinforcement learning and decision making, specifically focusing on exploration---the ability to deliberately gather information in order to evaluate alternatives and reduce uncertainty. We ask: Do contemporary large language model possess the capability to explore in-context?

In this talk, I will summarize our experiments toward answering this question. We focused on simple multi-armed bandit problems, three modern large language models (LLMs), and a variety of prompt designs and configurations. We found that only one configuration resulted in satisfactory exploratory behavior: GPT-4 with chain-of-thought reasoning and an externally summarized interaction history. All other configurations did not result in robust exploratory behavior. Thus we conclude that non-trivial algorithmic interventions may be required to empower LLM-based decision making agents in complex settings.

SD20

Summit - 340

Advancements in Online Sequential Decision-Making: Theory and Applications

Invited Session

Decision Analysis Society

Chair: Feng Zhu, MIT, Cambridge, MA, United States

1 - Actions as Instruments: An Online Strategic Decision-Making View

Rui Ai, MIT, Cambridge, MA, United States, David Simchi-Levi, Feng Zhu

It is common to observe only the equilibrium price and quantity in the market, while the entire demand curve remains inaccessible. This situation raises concerns about Ordinary Least Squares (OLS) leading to biased estimators due to confounding issues. We address this in a service fee pricing context by demonstrating a solution that utilizes proactive actions as instruments. Additionally, we illustrate that market randomness can be beneficial to some extent and identify the corresponding phase transition. Furthermore, considering agents with long-term perspectives, we develop a robust algorithm designed to counteract their strategic behaviors through low switching. Our novel algorithm achieves optimal regret upper bounds, which we compare to the associated lower bounds in various settings. Finally, we expand our approach to include generalized reinforcement learning, detailing the sample complexity involved in learning an ε-optimal policy.

2 - Generalization Analysis via Online Learning: New Bounds and Insights

Hao Liang, University of Illinois Urbana-Champaign, Champaign, IL, United States, Yunzong Xu

In this talk, we discuss an interesting framework for analyzing the generalization error of statistical learning algorithms via techniques from online learning theory. We present new upper and lower bounds and discuss their implications.

3 - Risk Detection and Inventory Coordination under Uncertain Time-To-Recover

PENGFENG SHU, Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Feng Zhu, Chung Piaw Teo, David Simchi-Levi

Time-To-Recover (TTR) is the duration needed for a supplier to restore full functionality after a disruption, serving as a fundamental metric for assessing supply chain risk exposure. Although previous studies typically presume complete knowledge of TTR, companies often possess only limited distributional information about TTR. In this work, we introduce a new supply chain resiliency framework that accommodates uncertain TTR for risk detection and inventory coordination across general supply chains. We propose a new Risk Exposure Index (REI) for pre-disruption risk assessment, which evaluates the expected lost sales using a prioritization model that underscores the necessity of prioritizing production of some products over others under uncertain TTR. Notably, using the new REI, the ranking between two nodes can completely “flip” in contrast to the conventional approach. We then theoretically analyze the cost of TTR uncertainty for intra-disruption inventory coordination by comparing the optimal prioritization decisions with the scenario-wise optimal solutions (those obtained when TTR is fully known). Our main result is a tight competitive ratio. Additionally, we extend our framework to include post-disruption inventory recovery, offering a broader view of the trade-offs between inventory recovery time minimization and lost sales minimization under uncertain

TTR. Through an extensive numerical case study, we illustrate the benefits of our new framework. Our results highlight that uncertain TTR can significantly alter the risk exposure of suppliers, emphasizing the importance of coordinated inventory management during a supply chain disruption.

4 - Rl for Ride-Hailing

Shuze Chen, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Tianyi Peng

This is a work about how reinforcement learning can facilitate online decision-making of large-scale ride-hailing system.

5 - Offline Oracle-Efficient Learning for Contextual MDPs via Layerwise Exploration-Exploitation Tradeoff

Haichen Hu, MIT, Cambridge, MA, United States, Jian Qian, David Simchi-Levi

In this paper, we address the stochastic Contextual Markov Decision Process (CMDP) problem with horizon H. We introduce a reduction from CMDPs to offline density estimation under the realizability assumption. We develop an efficient, statistically near-optimal algorithm requiring only O(H log T) calls to an offline density estimation algorithm (or oracle) across all T rounds of interaction. This number can be further reduced to O(H log log T) if T is known in advance. Our results mark the first efficient and near-optimal reduction from CMDPs to offline density estimation without imposing any structural assumptions on the model class. A notable feature of our algorithm is the design of a layerwise exploration-exploitation tradeoff tailored to address the layerwise structure of CMDPs.

SD21

Summit - 341

Control Applications in Finance

Invited Session

Decision Analysis Society

Chair: Ruijing Yang, Stevens Institute of Technology, Hoboken, NJ, United States

1 - Valuing Non-Myopic Views in Equity Premium Forecasting

Cheng Lu, Stevens Institute of Technology, Hoboken, NJ, United States, Majeed Simaan

There is mixed evidence on the relationship between macroeconomic variables and future stock returns. On the one hand, Welch and Goyal (2008) demonstrate that a naive moving average of stock returns outperforms individual macroeconomic variables in predicting the equity risk premium. On the other hand, a different stream of literature demonstrates that combining forecasts from macroeconomic variables enhances out-of-sample (OOS) predictions (Rapach et al., 2010). More recently, machine learning (ML) has shown a prominent ability to uncover hidden patterns in financial data. Given recent advances in artificial intelligence, ML stands as an appropriate candidate for predicting the equity premium. Nonetheless, one drawback of ML models is that they do not incorporate equilibrium characteristics and other forward-looking market mechanisms. In this regard, our paper fills the gap in the literature and proposes a non-myopic reinforcement learning (RL) approach that extracts equilibrium views about the equity risk premium, which we compare with 23 benchmarks used in the literature, including ML. While RL does not result in the highest OOS R^2, it provides the best performance in terms of Recall, i.e., identifying true market decline. This evidence is strongest during recession periods when hedging is needed the most. Our findings align with the nature of a risk-averse RL agent, underscoring the value of incorporating non-myopic views into forecasting models.

2 - A Global Discrete simulation Optimization Approach for Quantile Functions

Tianrui Zhao, Stony Brook University, Stony Brook, NY, United States

We propose a modification of the simulated annealing algorithm for solving discrete simulation optimization problems in which the objective function is the quantile function of a simulation output with noise. Unlike the original simulated annealing algorithm, our method uses a constant temperature, and it uses the number of visits made to different states to estimate the optimal solution. In contrast to performing multiple simulations at a candidate state to estimate its quantile performance, we propose an iterative framework that performs a single simulation for the candidate points encountered.

3 - Optimization in Energy Markets

Kai Chen, KPMG, New York City, NY, United States

Mathematical optimization techniques, such as linear programming and stochastic optimization, are commonly used to solve the complex optimization problems encountered in energy markets. This presentation delves into various facets of optimization within this dynamic domain, exploring its pivotal role in allocating resources and managing operations to maximize benefits or achieve specific objectives, such as valuation and risk management. First, we examine foundational market operations, focusing on power plant dispatching and the management of trading portfolios using Mixed Integer Linear Programming (MILP) algorithms. Subsequently, we illustrate the formulation for gas storage and pipeline systems with dynamic programming and a simplified version that uses only linear programming. Additionally, we explore the utilization of Battery Energy Storage Systems (BESS) in energy trading, outlining the requisite optimization techniques. Delving deeper, we navigate the intricacies of the Financial Transmission Right (FTR) market, recognized as one of the most complex arenas within energy markets, which often requires modeling the operations of the entire power grid. Finally, we present recent developments in the sector, offering insights into ongoing advancements.

4 - Optimal Portfolio Execution Strategies under Chance Constraints on Capital Adequacy Ratio Requirement

Ruijing Yang, Stevens Institute of Technology, Hoboken, NJ, United States, Zachary Feinstein, Somayeh Moazeni

We investigate optimal strategies to liquidate a portfolio of risky assets over a period of time under capital adequacy requirements subject to price impacts. Liquidating risky assets at every time step induces temporary impact on execution prices and permanent price impacts on the market prices. We study this problem with chance constraints on the regulatory requirements. Two versions of the chance constraints on the capital adequacy ratio requirements are provided: (i) require the capital adequacy requirement to be satisfied with a given probability at each individual time step or (ii) require the capital adequacy ratio time series to exceed the mandated level with a given probability. For the first problem, we develop a conservative approximation reformulation of the problem and establish sufficient conditions for its convexity. For the second problem, we use a stochastic approximation method to find the optimal liquidation strategy. Numerical studies are provided to study the optimal liquidation strategies in single-asset and multiple-illiquid asset cases. We show that in general the optimal portfolio execution strategy under the capital adequacy ratio requirement constraints differs from the optimal strategy in the absence of this requirement; in particular, when including these regulatory constraints, the liquidation strategy tends to postpone actions to the future. Computation costs of these two problems are also compared and we find that the conservative approximation method for the individual constraint problem is more computationally efficient compared to the stochastic approximation method for the joint constraint problem.

SD22

Summit - 342

Utility, Sharing, and the Value of Information

Invited Session

Decision Analysis Society

Chair: Andrea Hupman, University of Missouri - St. Louis, Saint Louis, MO, United States

1 - Risk Sharing with a TIME Preference

Ali Abbas, University of Southern California, Los Angeles, CA, United States, Zhengwei Sun

We consider the case of a partnership where individuals face an uncertain deal whose pay-offs will be realized at some future time. The individuals have (i) different risk tolerances and (ii) different time preference discount rates. Within this context, we derive the optimal share of each individual within the partnership. We discuss deviations and interpretations of the optimal shares from classic risk-sharing formulations that do not incorporate time preference.

2 - Wealth, Preferences, and the Value of Information

Andrea Hupman, University of Missouri - St. Louis, Saint Louis, MO, United States

This talk examines how characteristics of the decision maker affect the value of information (VOI) and how these effects have real-world implications. In particular, we consider how wealth effects and preferences of a decision maker influence the calculation of VOI. We characterize how simplifications in the formulation of VOI can create risk for the decision maker. We use the case of organizations purchasing information from a data broker to illustrate the implications of these results.

3 - Maximum Entropy Utility and Evaluation of Uav Designs

Robert Bordley, University of Michigan, Troy, MI, United States

In design problem, individuals make decisions on behalf of stakeholders. Stakeholder preferences are formalized as design targets. But assigning appropriate design targets is challenging. A maximum entropy utility is used to ensure that design decisions only reflect requirements and not other extraneous preferences. This paper derives a Gaussian utility function defined over the logarithm of a concave function. Certainty equivalents are easily computed when uncertainty is conjugate to the utility. The multivariate extension of the utility is also Gaussian with correlations describing complementarity and substitutability across target dimensions.

This utility can score how well different design solutions address different design targets.

4 - The Shapley Value of Information

Emanuele Borgonovo, Bocconi University, Milan, Italy, Francesca Beccacece, Emanuele Borgonovo, Giovanni Rabitti, Andrei Savochkin

In this work, we present the notion of Shapley value of information. In a 1988 Risk Analysis paper, Howard proposes a decomposition of the information value across sources of uncertainty. The decomposition, however, depends on the order with which sources of information are considered. We propose a Shapley decomposition that uniquely apportion the total information value in a decision problem to each information source. We prove several properties of the decomposition and address computational issues.

5 - Discounted Incremental Utility

Manel Baucells, Darden School of Business, Charlottesville, VA, United States, Michał Lewandowski, Krzysztof Kontek

Most decisions involve multiple payoffs over time and under risk, as for instance the sequential play of a lottery. For an individual who cares not only about profit at the end, but also on how early these profits accrue, we apply a modified version of the axioms of subjective expected utility to obtain the Discounted Incremental Utility model. To this classical bedrock, we add the notion that preferences are affected by range effects. This modification results in the so-called Range-Discounted Incremental Utility, which can successfully predict a plethora of phenomena such as attitudes towards sequential play, the four-fold pattern, preference for temporal hedging, preference reversals for risk and time, and patterns of decreasing impatience. For comparison, we also introduce a rank-dependent modification and examine its shortcomings.

SD23

Summit - 343

Multiple Objective Optimization and Decision Processes

Invited Session

Multi Criteria Decision Making

Chair: Honggang Wang, California State Poly Tech Pomona, 3801 West Temple Avenue, Pomona, 91768, United States

1 - Continuation Search in Manifold Surface for Multi-Objective Optimization

Honggang Wang, Cal Poly Pomona, Pomona, CA, United States

Imagine one day, your breathing became consistently labored and shallow. Months later you were finally diagnosed with pulmonary fibrosis, a disorder with no known cause and no known cure, created by scarring of the lungs. If that happened to you, you would want to know your prognosis. That’s where a troubling disease becomes frightening for the patient: outcomes can range from long-term stability to rapid deterioration, but doctors aren’t easily able to tell where an individual may fall on that spectrum. Machine learning analytics may be able to aid in this prediction, which would dramatically help both patients and clinicians. In this research, we will predict a patient’s severity of decline in lung function based on a CT scan of their lungs together with some other patient clinic data. The lung function is determined based on FVC output from a spirometer, which measures the volume of air inhaled and exhaled. One major challenge of this work is about accurately uncovering such multiple-dimensional temporal patterns unstructured 3D scan images and time series. Some preliminary results using real-life patient datasets will be presented and discussed.

2 - Integrating Simulation and Multiple-objective Optimization Using MGD-SPSA Algorithm

Haobin Li, National University of Singapore, Singapore, Singapore, Yan Li, Ying Huang , Ek Peng Chew

Constructing a Common Descent Vector (CDV) from a combination of gradients offers a practical method to address Multi-Objective Optimization (MOO) challenges by simultaneously minimizing multiple objectives. This approach, however, faces difficulties in environments where gradients are unavailable, such as in black-box simulations. To overcome this, our paper presents the Multiple Gradient Descent-Simultaneous Perturbation Stochastic Approximation (MGD-SPSA) algorithm. This advanced algorithm utilizes biased gradient estimations from the SPSA to construct an estimator for the CDV, aiming for Pareto optimality. Leveraging the characteristics of the SPSA, which perturbs all decision variables at once and approximates the gradient of the objective function using only two function evaluations per iteration, the MGD-SPSA achieves substantial computational efficiency in high-dimensional settings. We detail its algorithmic process for both bi-objective and many-objective scenarios and provide theoretical evidence of its strong convergence. The effectiveness of MGD-SPSA is demonstrated through ZDT test functions and diverse simulation-optimization frameworks, showing significant reductions in CPU time and enhancements in the quality of Pareto solutions compared to traditional evolutionary algorithms. These findings highlight the transformative impact of integrating advanced simulations with MOO strategies, facilitating optimal trade-offs across various objectives in different industrial systems.

3 - A new approach for Selection Green Supplier Problem: A Q-Rof MCDM Approach

Babek Erdebilli, AYBU, Ankara, Turkey, Sara Saberi

The selection of green suppliers (GSS) has emerged as a crucial aspect in sustainability initiatives, focusing on efficient natural resource utilization. In a case study pertaining and selection are conducted using the intuitionistic Q-Rof MCDM method, which accommodates both qualitative and quantitative criteria. Introducing a novel perspective on the GSS, this innovative approach offers a fresh methodology for selecting environmentally responsible suppliers. In response to the growing importance of sustainability in supply chain management, this approach addresses the need for efficient and effective green supplier selection processes. By integrating advanced decision-making techniques and environmental criteria, this method aims to optimize supplier selection while promoting eco-friendly practices throughout the supply chain. environmental applications, and recycling. Q-Rof MCDM is employed due to its suitability for evaluating decision-makers and criteria. The hybrid method combining Q-Rof and MCDM proves highly effective in selecting the most suitable supplier among alternatives, offering potential integration into similar decision-making scenarios.

4 - PRIORITIZING PATIENT SAFETY ISSUES: A MULTI-CRITERIA APPROACH USING MEREC-CoCoSo INTEGRATION

Fatma Pakdil, Eastern Connecticut State University, Willimantic, CT, United States, Pelin Toktas, Gulin Feryal Can

Patient safety is one of the highly prioritized performance areas in the healthcare systems. Despite healthcare facilities and practitioners put a substantial effort on implementing patient safety improvements, there is still an uncertainty about prioritizing patient safety needs. In this context, healthcare organizations are expected to follow a systematic prioritization strategy for potential patient safety projects to allocate their scarce resources more effectively in order to achieve zero harm to patients. Complex nature of patient safety issues necessitates a well-structured prioritization and selection process to allocate scarce resources more effectively and efficiently onto patient safety improvement efforts. Given the complexity and urgency of such decision-making processes, decision-makers need multi-criteria decision-making (MCDM) methods to prioritize and select potential patient safety issues in a given time. In light of this, this study aims to exemplify how MCDM methods can be employed in patient safety issue prioritization and selection processes in acute care hospitals. This study proposes a model that implements the Method based on the Removal Effects of Criteria (MEREC)-Combined Compromise Solution (CoCoSo) method to show the practicality and feasibility of MCDM methods for such needs. The proposed model employs MEREC to evaluate and weight the criteria and CoCoSo to prioritize the alternative patient safety issues. This is the first study that employs an MCDM approach to prioritize and select patient safety issues in the healthcare industry.

5 - Multi-objective cooperative game theory for improving coordination in humanitarian operations

Ayesha Farooq, Kansas State University, Manhattan, KS, United States, Jessica Heier Stamm

In humanitarian settings, coordination is pivotal for efficient and effective operations given the presence of diverse decision makers and challenging conditions. We integrate cooperative game theory and optimization to enhance multi-objective, multi-stakeholder coordination. Our approach fairly allocates system-wide costs and utilities among all participating organizations. The model addresses the inherent tension

between demand fulfillment and cost efficiency by applying Chebyshev goal programming. The model minimizes the maximum relative deviation from predefined goals, ensuring rationality at both individual and collective levels. We analyze player behavior under several policies for allocating the costs and benefits of coordination. When stakeholders' perspectives are considered, organizations can identify unified strategies that maximize their collective impact.

6 - Multi-Objective Optimization in Decision Spaces

Kalyanmoy Deb, Michigan State University, East Lansing, MI, United States

Evolutionary multi-objective optimization (EMO) algorithms are capable of finding multiple near Pareto-optimal (PO) solutions simultaneously in a single run for two to about 20-objective problems. These algorithms use a population of solutions in each iteration and update the population by selecting better solutions using domination and diversity-preserving techniques and then by mixing them to create new and hopefully better solutions. While the obtained PO solutions are expected to be well distributed in the objective space, they may not be well-distributed in a decision space which a decision-maker (DM) may prefer. For example, if a relative importance of objectives are known a priori, the DM may be interested in solutions well distributed in an importance metric space clearly indicating a hierarchy of PO solutions from highest to the lowest important objectives, rather than having a set of PO solutions with a good distribution in the objective space. In this talk, we introduce a number of decision spaces which a DM may prefer in finding a good distribution of solutions in that space and then present three different EMO approaches to find a preferred distribution of PO solutions using modified EMO approaches. The advantages and disadvantages of these approaches will be highlighted by applying them on a number of test and engineering design problems. This study makes the EMO approaches more generic and bring the importance of decision-making aspects within the task of multi-objective optimization.

SD24

Summit - 344

Optimizing Impact in Nonprofits and Alliances

Contributed Session

Contributed

Chair: Ye Eun Choi, Graduate School of Data Science, Gwanakro 1 gil, Gwanakgu, Seoul, South Korea, Seoul, 08826

1 - Targeting broadcasts to moderate opinion dynamics in social networks

Philip Solimine, University of British Columbia, Vancouver, BC, Canada, Wei Li

We discuss the solution to a restricted or coarse targeting problem in social networks. In this problem a benevolent social planner moderates opinion dynamics in a network of naive agents, by broadcasting a common signal to each node with varying intensity. For a class of natural quadratic planner objectives, the coarse targeting problem has a unique solution that is amenable to numerical optimization. Generally the vector of optimal broadcast strengths is a combination of network centrality with the final state of the autonomous network. We explore moderation and stability of opinions under various moderation objectives and find that the planner is frequently incentivized to broadcast polarizing signals, even when her goal is to synchronize opinions. We also characterize the optimal target weights in a number of special cases, in which symmetries in the network topology allow for the solution to the restricted targeting problem to be characterized in terms of a small set of moments of the distribution of broadcast weights through the network.

2 - An optimization framework to provide volunteers with task selection autonomy and group opportunities on nonprofit crowdsourcing platforms

Milan Kaur, Rensselaer Polytechnic Institute, Troy, NY, United States, Jennifer Pazour

Nonprofit Organizations (NPOs) rely on volunteers to support community needs but due to lack of available volunteer information, can seldom make strategic task assignments to achieve successful volunteer participation and NPO task completion. Making successful volunteer group-to-task assignments is challenging because (i) there are exponentially many ways a set of volunteers can be assigned in groups; and (ii) NPOs tend to have limited data concerning volunteers' personal preferences, availability, and skills. To address these challenges, we introduce a framework that integrates a set of static integer programs that enables nonprofits to offer volunteers a menu of tasks to choose from and creates ideal volunteer groups assigned to tasks. The menu creation (MC) IP offers volunteers with personalized task recommendations menus. The group creation (GC) IP creates homogenous volunteer groups with ideal learning potential and affinity among volunteers to encourage their motivations to participate through successful engagement, networking, skill learning and utilization. Our methodology is implemented with a case study representing a partner NPO with online tasks. Results indicate that the implementation of the MC methodology provides a statistically significant improvement to nonprofits and volunteers. Implementing the GC methodology helps obtain a statistically significant increase in ideal group creations but results in a tradeoff of decrease in other NPO KPIs. This static model is being extended to a dynamic integer linear program with a sequential decision-making framework. The aim of the dynamic model is to identify the impact of assignment decisions on task completion and volunteer retention in future periods.

3 - Mapping the past to Build the Future: a Visualization System for Enhancing Humanitarian Aid Effectiveness Through Project Topic Modeling Based on Final Evaluation Reports

Yeeun Choi, Seoul National University, Seoul, Korea, Republic of, Jinwoo Choi, Hyunwoo Park

This study aims to support the more effective planning and execution of new development cooperation projects by visualizing the relationships among completed projects based on specific criteria. As the number of aid organizations continues to increase and the overall number of projects expands, it has become challenging to identify which aid projects were carried out. This difficulty leads to overlapping projects and repeated mistakes, resulting in reduced aid effectiveness. To address these issues, we developed a system to support various actors in development cooperation by providing comprehensive visualizations of project networks. The system employs final evaluation reports of projects conducted by various organizations and visualizes the similarity network of completed projects based on these reports. To allow users to search for the desired type of project based on specific criteria, we used a predefined topic modeling method. First, we established criteria such as the Sustainable Development Goals (SDGs), regions, and cultural areas, and defined detailed classes to characterize projects within each criterion. Then, extracted keywords for each class were used for predefined topic modeling method to quantify the characteristics of each project based on these criteria. Through this, the system visualizes the project network, allowing users to intuitively identify similarities among previously implemented projects and offering links to download each project's final evaluation report. By facilitating the review of past project outcomes, limitations, and lessons learned, the system aims to enhance the planning and execution of more effective aid projects by leveraging empirical knowledge gained from similar past initiatives.

SD25

Summit - 345

Predictive Models and Strategies in Global Health

Contributed Session

Contributed

Chair: Robin Na

1 - Does Reading the Literature Improve the Ability to Predict the Effects of Interventions?

Robin Na, Massachusetts Institute of Technology, Cambridge, MA, United States, Abdullah Almaatouq

Making effective policy recommendations remains notoriously challenging due to the formidable heterogeneity and contingency we must navigate to predict the outcomes of interventions. While science is expected to generate cumulative knowledge that helps people understand such complexity, this raises an empirical question of which set of academic papers could shift human priors in a direction that enhances predictions. Motivated by the infeasibility of conducting controlled human-subject experiments of such a nature, we propose a novel method for investigating this fundamental question by using large language models (LLMs) and retrieval-augmented generation (RAG) as proxies for exploring the treatment effect of research papers. Each version of GPT-4, augmented with a specific set of behavioral research papers, is asked to predict intervention outcomes of cooperation under the high-dimensional, integrative experiment design space of a public goods game (PGG). The space consists of 211 unique experiment designs with 20 dimensions of varying conditions to map the rich heterogeneity of treatment effects. We demonstrate the LLM’s ability to retrieve academic papers and alter its distribution of predictions for experiment outcomes in directions that are expected based on the documents' contents. However, such shifts do not contribute to higher accuracy for the most time and our method explores the features of papers that improves such accuracy. The framework introduces a method for evaluating the potential contribution and informativeness of scientific literature in prediction tasks, while also showing how our integrative experiment can be used as a benchmark for LLM's ability to make predictions about human behavior.

2 - Strategies for the Global Alliance for Vaccines and Immunization in International Vaccine Pricing and Procurement

Banafsheh Behzad, California State University, Long Beach, Long Beach, CA, United States, Susan Martonosi, Ruben Proano, Zooey Meznarich

The Global Alliance for Vaccines and Immunization (Gavi) is an organization that works with manufacturers and governments to mediate the pricing and procurement of vaccine products and facilitates the transfer of technology to manufacturers. Their work with UNICEF supports low-income countries by improving access to pediatric vaccines. Multivalent vaccines, which inoculate against multiple diseases, improve efficiency of immunization. However, the production of multivalent vaccines is often limited, as is the case for the hexavalent (6-antigen) pediatric vaccine. Because Gavi negotiates large contracts on behalf of many countries, they can incentivize manufacturers to produce hexavalent. Existing models of vaccine markets with price negotiations between external organizations and private manufacturers are adapted and applied to the international vaccine market. Mathematical programming tools are used to examine optimal prices and quantities within the global vaccine market, while incentivizing the production of multivalent vaccines. As a result, relationships among manufacturer capacity, demand projections, and price provide insight into future policy recommendations and feasibility.

3 - A Stochastic Optimization Model for Scheduling Tenders for Pediatric Vaccines in Low- and Low-Middle-Income Countries

Nicholas Uhorchak, University of Arkansas, Fayetteville, AR, United States, Ruben Proano, Sandra Eksioglu, Fatih Cengil, Burak Eksioglu

Pediatric vaccine tender scheduling represents a strategic approach for countries to efficiently plan and procure essential vaccines over extended periods. This process, facilitated by organizations like UNICEF (United Nations Children’s Fund), PAHO (Pan American Health Organization), and GAVI (Gavi, the Vaccine Alliance), aims to ensure fairness and effectiveness in vaccine distribution. However, it's a multifaceted challenge, balancing the need to meet demand, maintain an ample supply, and navigate logistical hurdles. Complicating matters further are the unpredictable fluctuations in vaccine demand and manufacturing capacity. To tackle these complexities, we propose a stochastic optimization model. This approach dynamically generates vaccine tender schedules, taking into account uncertainties, thus enabling proactive planning over long timeframes. Our model addresses key research questions: How can we optimize the sequencing and timing of vaccine tenders to maximize affordability and profit? What is the ideal procurement schedule for single or multiple antigen scenarios? And how do uncertainties in demand and manufacturing capacity influence tender scheduling? Leveraging real-world data and forecasting techniques, we test our model's efficacy. Our analysis reveals that our approach yields robust first stage (i.e., here and now) decisions and adeptly adapts to stochastic factors in the second stage (i.e., wait and see) decisions. We demonstrate the feasibility of generating optimal procurement schedules for various scenarios, while effectively responding to fluctuations in stochastic inputs.

4 - Resilience strategies for epidemic testing capacity: pooling, scheduling, and flexible machine allocation

Erica Gralla, George Washington University, Washington, DC, United States, Nadia Lahrichi, Jad El Hage, Fannie Côté

The COVID-19 pandemic brought to light a critical problem in the global pandemic response infrastructure: insufficient testing capacity. Major epidemics stress existing systems with a surge in demand for testing, leading to very long turnaround times for results. This paper identifies and explores sources of fragility in testing laboratories and develops strategies for resilience, including pooling specimens, scheduling personnel, and flexibly allocating liquid handling machines. These resilience strategies are evaluating using processes and data from two diverse COVID-19 testing facilities: a Nepalese laboratory system relying heavily on manual operations, and a US university-based laboratory relying on automation. Discrete event simulations and a scheduling algorithm test the impact of the resilience strategies and identify when and why they are beneficial. The results show that pooling strategies are beneficial for meeting surges in demand, but require more liquid handling capacity. Flexibility in personnel scheduling and machine allocation also increases the laboratory capacity in centers that rely heavily on manual or automated tasks, respectively. With these strategies, testing facilities could increase their resilience to surges in testing demand by better leveraging their current resources without significant additional investment.

SD26

Summit - 346

Blockchain Applications Across Industries

Contributed Session

Contributed

Chair: Mengfan Xu, Northwestern University, 2145 Sheridan Road, Evanston, IL, 60208, United States

1 - An efficient node selection and validation method for providing decentralization in a blockchain with multiple dynamic consensus groups

Jin Choi, Ajou University, Suwon, Korea, Republic of, Jon Hwi Lee, Jaeyoung Lee

In a blockchain network with multiple dynamic consensus groups, it is very important to provide decentralization by designing an efficient method for node selection and validation for each group. Especially, we consider decentralization from two points of views such as node point and stake point. Even for the stake point of view, it is not easy to attain decentralization because any node investing large stakes might have big impact on generating blocks by participating in the consensus process. In this work, we devise a good policy for providing decentralization for the considered cases by performing a numerical experiment, while considering some performance measures such as fairness, entropy, and gini coefficient.

2 - Competitiveness and Cross-Ownership of Startups in the Blockchain Industry: a Network-Based Empirical Analysis

Samaneh Samiee, University of Texas at Arlington, Arlington, TX, United States

This study investigates two key challenges of running an early stage blockchain based business in different markets. It examines whether the competitiveness of the field of activity of a blockchain startup affects its success. Moreover, it studies the (dis)advantages of accepting venture capital funding from investors who own other blockchain startups in their portfolios. Additionally, this study examines whether the competitiveness of the startup's field of activity and cross-ownership moderate the relationship between its entry to market and its success and survival. Utilizing a Bipartite network analysis approach, we develop an empirical model by analyzing network-based relations between startups founded between 2008 and 2023. We find empirical evidence for the moderating roles of competitiveness and cross-ownership. Thus, the results show that the advantages of choosing a competitive field of activity and being cross-owned with other blockchain startups outweigh the business stealing anti-competitive effects. The managerial and theoretical implications of this research brings the ongoing entrepreneurship debates on these scenarios to the technology entrepreneurship for the first time and establishes how business applications of blockchain moderates business success.

3 - The Influence of Blockchain Traceability on Consumer Purchase Intention in The Organic Food Industry

Jin-Jin Lyu, Kyungpook National University Business Administration, Daegu, Korea, Republic of, Chae-Bogk Kim

Embracing the transformative potential of blockchain traceability label enables enterprises to carve out new avenues for consumer engagement and brand differentiation, ultimately fostering sustainable growth in a fiercely competitive market environment. This study innovatively explores the impact of blockchain traceability label on consumer purchase intention within the organic food industry. A total of 975 valid questionnaires were collected online via a paid survey platform, and one-way ANOVA was employed to ascertain the significance of label variations on purchase intention. Additionally, we utilized two-way ANOVA to unveil the intricate interplay between brand maturity, income level, and the influence of label on purchase intention. The analysis demonstrates that consumers exposed to blockchain traceability label exhibit significantly higher purchase intention compared to those encountering standard or no label. Furthermore, it was observed that high-income consumers exhibit heightened purchase intention when presented with products featuring blockchain traceability label. Moreover, the presence of blockchain traceability label on product packaging disrupts traditional brand preference theories by amplifying purchase intention for new brand. These innovative findings offer valuable insights for enterprises seeking to navigate the evolving consumer landscape and stay ahead of the curve.

4 - The Impact of Blockchain Adoption on Vertically Differentiated Competition

Yuhan Yan, Renmin University of China, Beijing, China, People's Republic of, Xiaofang Wang

The rapid proliferation of blockchain technology has changed the management practice of firms in many aspects. This paper considers pricing competition between two firms offering vertically differentiated products, as well as the quality improvement decision of the low-quality firm. Our paper captures two roles blockchain can play in practice: transmitting reliable quality signals and expanding sales toward technology-sensitive customers. Our results highlight the interactive impacts of blockchain adoption on quality-differentiated firms under different decision sequences.

5 - Decentralized Blockchain-based Robust Multi-agent Multi-armed Bandit

Mengfan Xu, Northwestern University, Evanston, IL, United States, Diego Klabjan

We study a robust multi-agent multi-armed bandit problem where multiple clients or participants are distributed on a fully decentralized blockchain, with the possibility of some being malicious. The rewards of arms are homogeneous among the clients, following time-invariant stochastic distributions that are revealed to the participants only when the system is secure enough. The system's objective is to efficiently ensure the cumulative rewards gained by the honest participants. To this end and to the best of our knowledge, we are the first to incorporate advanced techniques from blockchains, as well as novel mechanisms, into the system to design optimal strategies for honest participants. This allows various malicious behaviors and the maintenance of participant privacy. More specifically, we randomly select a pool of validators who have access to all participants, design a brand-new consensus mechanism based on digital signatures for these validators, invent a UCB-based strategy that requires less information from participants through secure multi-party computation, and design the chain-participant interaction and an incentive mechanism to encourage participants' participation. Notably, we are the first to prove the theoretical guarantee of the proposed algorithms by regret analyses in the context of optimality in blockchains. Unlike existing work that integrates blockchains with learning problems such as federated learning which mainly focuses on numerical optimality, we demonstrate that the regret of honest participants is upper bounded by $\log{T}$. This is consistent with the multi-agent multi-armed bandit problem without malicious participants and the robust multi-agent multi-armed bandit problem with purely Byzantine attacks.

SD27

Summit - 347

Current Research on AI

Invited Session

New Product Development

Chair: Philipp Cornelius, Rotterdam School of Management, Rotterdam, Netherlands

1 - Generative AI at School: How ChatGPT Changes Student Learning

Philipp Cornelius, Rotterdam School of Management, Rotterdam, Netherlands, Matthias Lehmann, Fabian Sting

Generative AI tools such as ChatGPT are transforming education. At any hour of the day, for any subject, students now have access to a personalised AI tutor, who can answer questions and explain difficult concepts. However, students may also be tempted to rely on AI for more than tutoring. Many educational programmes rely substantially on homework assignments for practice and assessment. If students use generative AI to reduce the amount of time spent on homework, they may engage less with educational material and in effect learn less. In this paper, we study how the use of generative AI during homework assignments affects students’ short-term performance in the assignments as well as their long-term learning.

2 - Artificial Intelligence, Ceo Turnover, and Exploration Orientation in Firm Innovation

Xinyu Ma, University of Pennsylvania, Philadelphia, PA, United States, Bowen Lou, Lynn Wu

This study examines the role of artificial intelligence (AI) in facilitating corporate innovation following a CEO turnover event. Using patent data for firms that have undergone a CEO turnover, we find that firms with greater AI investment are more successful in explorative innovation. Further analyses show that this effect is likely due to the firms with AI investment being better positioned to enable strategic change in innovation, cultivate a culture of exploring frontiers in innovation, and manage R&D. Importantly, AI mitigates myopic management of CEOs and overcomes the barrier of information overload that they often encounter. A new CEO can also direct more resources to acquiring employees with AI skills to facilitate explorative innovation. Overall, this study highlights the significant value of AI in promoting exploration in innovation during organizational transition periods. By shedding light on the role of AI in guiding innovation and fostering change, we provide insights for AI-driven transformation and creativity in organizations when navigating dynamic environments.

3 - Generative AI and Copyright: A Dynamic Perspective

Song Yang, London Business School, London, United Kingdom

The rapid advancement of generative AI is poised to disrupt the creative industry. Amidst the immense excitement for this new technology, its future development and applications in the creative industry hinge crucially upon two copyright issues: 1) the compensation to creators whose content has been used to train generative AI models (the fair use standard); and 2) the eligibility of AI-generated content for copyright protection (AI-copyrightability). In this paper, we aim to better understand the economic implications of these two regulatory issues and their interactions. By constructing a dynamic model with endogenous content creation and AI model development, we unravel the impacts of the fair use standard and AI-copyrightability on AI development, AI company profit, creators’ income, and consumer welfare, and how these impacts are influenced by various economic and operational factors. For example, while generous fair use (use data for AI training without compensating the creator) benefits all parties when abundant training data exists, it can hurt creators and consumers when such data is scarce. Similarly, stronger AI-copyrightability (AI content enjoys more copyright protection) could hinder AI development and reduce social welfare. Our analysis also highlights the complex interplay between these two copyright issues. For instance, when existing training data is scarce, generous fair use may be preferred only when AI-copyrightability is weak. Our findings underscore the need for policymakers to embrace a dynamic, context-specific approach in making regulatory decisions and provide insights for business leaders navigating the complexities of the global regulatory environment.

4 - Examining the Impact of Generative AI on Users’ Questioning Behavior: Evidence from a Natural Experiment of Pilot Usage of ChatGPT

xinzhi rao, University of Florida Warrington College of Business Administration, Gainesville, FL, United States, Guohou Shan, Liangfei Qiu

Our study examines the impact of the usage of generative AI on users’ questioning behavior on Stack Overflow, one of the largest Q&A platforms. Leveraging a natural experiment of piloting the usage of ChatGPT during its launch and ban on the platform, we employ the difference-in-differences (DID) estimation to investigate the effects of generative AI on the quantity and quality of users’ questioning behavior, measured by the number of questions generated per week, question length, novelty, and received upvotes. Our findings reveal that using generative AI increases the number of questions generated by users. Additionally, these questions tend to be longer in length, more novel in content, and receive more upvotes. This suggests that users treat generative AI as an assistant, helping them generate higher-quality questions at a faster speed. We also explore the moderating effects of the number of answers generated (proxy as an experience) and the users' received badge (proxy as the reputation or celebrity on the platform) on the impact of generative AI to test the mechanisms. The implications of this study are both theoretical and practical. Theoretically, we contribute to the Information Systems literature by examining the impact of generative AI on users' questioning behavior in the context of Q&A platforms. Practically, our findings provide knowledge platform owners and managers with a better understanding of how generative AI influences users' knowledge contribution behavior in terms of asking questions. This understanding can guide decision-making and strategy development regarding the integration of generative AI on their platforms.

SD28

Summit - 348

Business and Climate Change

Invited Session

MSOM: Sustainable Operations

Chair: Safak Yucel, Georgetown University, Washington, DC, United States

Co-Chair: Soudipta Chakraborty, University of Kansas, Lawrence, 66213, United States

1 - How Does Flexibility Affect Emissions Performance?: Evidence from U.S. Power Generation

David Drake, Leeds School of Business, University of Colorado, Boulder, CO, United States, Suresh Muthulingam

Several CO2 abatement paths alter the flexibility burden placed on the conventional power generation grid. However, the environmental consequences of this flexibility are unexplored. In this study, we examine the environmental impact of flexibility in US power generation, distinguishing between the environmental effects of possessing flexibility and exercising flexibility.

2 - Additionality of Carbon Offsets: Project-Specific vs. Standardized Baselines

Soudipta Chakraborty, University of Kansas, Lawrence, KS, United States, Vishal Agrawal, Safak Yucel

A carbon offset represents reduction in emissions that can be used to compensate for emissions elsewhere. Developers generate offsets through emissions-reduction projects. A non-profit carbon registry issues offsets to developers while ensuring *additionality:* An offset needs to represent one unit of reduction from a developer's business-as-usual emissions, i.e., what the developer's emissions would have been in the absence of the market for offsets. Accordingly, environmental groups raise greenwashing concerns against non-additional offsets. Ensuring additionality is challenging because business-as-usual emissions is a developer's private information and an unobservable counterfactual for the registry. In practice, the registry assigns a baseline to represent business-as-usual emissions through one of the two methods: Under the *project-specific* method, a developer self-reports its business-as-usual emissions to the registry, which then inspects the report and assigns reported emissions as the baseline if it accepts the project. Under the *standardized* method, the registry assigns a common baseline to a group of developers. We compare economic and environmental implications of these two methods.

3 - The Implications of Minimum Recycled Content Requirements on Product Design and Recycling Channel Decisions

Donghyun Choi, Georgia Institute of Technology, Atlanta, GA, United States, Andre Calmon, Beril Toktay

Extended producer responsibility (EPR) is increasingly implemented across various industries through minimum recycled content requirements, presenting significant operational challenges for producers in securing recycled materials. Concurrently, producers are adopting materials with lower carbon footprints to adapt to a carbon-constrained business environment. In this context, we examine whether regulations promoting circularity synergize or conflict with carbon reduction targets. To address this question, we develop a stylized model to analyze a producer's product design and recycling channel decisions to comply with minimum recycled content requirements, where producers can either purchase recycled content (open-loop) or internally recycle end-of-use products (closed-loop). Our findings indicate that producers may face trade-offs between pursuing material circularity and carbon reduction under current regulations or voluntary carbon reduction targets, which mainly account for the firm's internal operations (i.e., scope 1 and 2). We suggest that incorporating supply chain emissions (scope 3) into regulations or carbon targets can promote both material circularity and carbon reduction. This study provides managerial insights on managing and identifying synergies among multiple environmental performance metrics and offers policy implications to further foster green innovation in product and supply chain management.

SD29

Summit - 420

Network Optimization in Power Systems

Invited Session

OPT: Network Optimization

Chair: Jim Ostrowski, University of Tennessee, Knoxville, TN, United States

Co-Chair: Jian Li, Stony Brook University, Stony Brook, NY, United States

1 - Strategic DG Placement and Sizing for Microgrid Formation using Sample Average Approximation

Dhiraj Pokhrel, University of Tennessee, Knoxville, Knoxville, TN, United States, Hugh Medal

Severe blackout events due to extreme weather are increasing in the recent years. Therefore, highlighting the need of resilient power distribution networks. Microgrid formation through strategic distributed generation (DG) placement is one of the measures that can be done to increase distribution system resilience. In this research we have formulated the DG placement and sizing problem as the two-stage stochastic mixed-integer program. The first stage decisions are DG placement and sizing decisions while the second stage is the network restoration decisions, which include which switches to close and which existing lines to open. The objective is to minimize the cost associated with the load shed and power generation. We tested our model in IEEE 33 bus system. Because the formulated stochastic problem contains extremely large number of scenarios, we used a sampling method to solve it. We present a detail computational study of the problem using a sample average approximation (SAA) method. Computational results on IEEE test case network demonstrate the effectiveness of the SAA method.

2 - Optimizing Discretized Time Network Train Scheduling: Introducing Adaptive Valid Inequalities and Computational Insights

Masoud Barah, Northwestern University, Evanston, IL, United States

Time-discretized train scheduling problems suffer from the issue of near-identical sub-optimal solutions that significantly slow down exact solution algorithms, especially when the fidelity of time-discretization is higher. One efficient method to eliminate these solutions is by lifting the solution space to an aggregated space, wherein the set of near-identical solutions is mapped into new variables. To effectively utilize the aggregated space, this paper proposes a set of minimal cover inequalities that tighten the lifted polyhedron of the aggregated space. Additionally, we present a heuristic algorithm based on the concept of arbitrarily extending lengths of arcs while dynamically aggregating the original space. Computational experiments demonstrate a 44% speedup in solution times and a reduction in the optimality gap when employing our valid inequalities and the heuristic.

3 - Structured Learning and Optimization: Foundations and Practice

Jian Li, Stony Brook University, Stony Brook, NY, United States

Next-generation (NextG) networked systems enabling applications ranging from 5G/6G, transportation, smart cities to autonomous driving will be data-driven and large-scale. This poses enormous burdens on traditional systems such as wireless networks, and emerging architectures such as edge and cloud computing systems, in which classic algorithmic guarantees do not always hold and rigorous performance evaluation can be even difficult. Though data-driven techniques have been applied to a variety of networking and distributed systems, pure data-driven solutions face their own challenges on curse-of-dimensionality and generalization. In this talk, I will present a structured reinforcement learning framework for NextG networked systems, which overcomes these challenges. This framework exploits the inherent structure encoded in classical models, and enables the design of novel learning architectures with low-complexity and improved sample efficiency.

4 - Optimizing Transport of Valuables: A Multi-Objective Approach Under Uncertainty

Fatemeh Zandieh, Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden, Alireza Fallahtafti, Armin Hamedizad

Transporting valuables and cash entails inherent risks, prompting the integration of security measures into routing strategies. This study presents a new approach, the multi-objective valuables/cash routing problem (MO-VCRP), to efficiently manage transportation. Our model aims to minimize both transportation risks and travel costs concurrently. To address inherent uncertainty, we propose utilizing an interval type-2 fuzzy controller to estimate risk, considering several factors. Additionally, we introduce a type-2 fuzzy multi-objective genetic algorithm (T2F-MOGA) to optimize MO-VCRP.

SD30

Summit - 421

Operator Splitting Methods for Optimization

Invited Session

OPT: Computational Optimization and Software

Chair: Jonathan Eckstein, Rutgers University, Piscataway, NJ, 08854, United States

1 - A Practical and Optimal First-Order Method for Large-Scale Convex Quadratic Programming

Jinwen Yang, University of Chicago, Chicago, IL, United States, Haihao Lu

Convex quadratic programming (QP) is an important class of optimization problem with wide applications in practice. The classic QP solvers are based on either simplex or barrier method, both of which suffer from the scalability issue because their computational bottleneck is solving linear equations. In this paper, we design and analyze a first-order method for QP, called restarted accelerated primal-dual hybrid gradient (rAPDHG), whose computational bottleneck is matrix-vector multiplication. We show that rAPDHG has a linear convergence rate to an optimal solution when solving QP, and the obtained linear rate is optimal among a wide class of primal-dual methods. Furthermore, we connect the linear rate with a sharpness constant of the KKT system of QP, which is a standard quantity to measure the hardness of a continuous optimization problem. Numerical experiments demonstrate that both restarts and acceleration can significantly improve the performance of the algorithm. Lastly, we present PDQP.jl, an open-source solver based on rAPDHG that can be run on both GPU and CPU. With a numerical comparison with SCS and OSQP on standard QP benchmark sets and large-scale synthetic QP instances, we demonstrate the effectiveness of rAPDHG for solving QP.

2 - Breaking the Cycle: Deterministic Block-Iterative Analysis for the Frank-Wolfe Algorithm

Zev Woodstock, James Madison University, Harrisonburg, VA, United States, Gabor Braun, Sebastian Pokutta

Over the last decade, Block-Coordinate Frank-Wolfe (BCFW) algorithms have led to the advent of many efficient first-order algorithms with applications in ML and engineering. However, to-date, they are only proven to work under restrictive assumptions on how the block-coordinates are updated, namely, (A) single-coordinate updates (via stochastic and greedy selection), and (B) cyclic/full update schemes which require the evaluation of all LMOs in the Cartesian product for every outer iteration. These settings are particularly disadvantageous when the coordinate LMOs require significantly different amounts of time (e.g., when some LMOs are cheap, while others require an eigencomputation or a large LP solve). We prove that the BCFW algorithm converges with a much milder "block-iterative" assumption, allowing for (I) progress without activating the most-expensive LMO(s) at every iteration, and (II) more flexibility in parallelized updates. We prove that our results are valid for several stepsize strategies, and we demonstrate a concrete gain in speed using *FrankWolfe.jl*.

3 - DEFBAL Part I: A Connection Between the ADMM and Proximal Gradient Methods

Jonathan Eckstein, Rutgers University, Piscataway, NJ, United States, Chang Yu, Paulo J. S. Silva, Felipe Atenas

The two minimization steps of the ADMM can be interpreted as the proximal gradient method applied to a particular dual formulation of the augmented Lagrangian subproblem. This talk explains this connection and how it can be used to create new ADMM-like convex optimization methods by substituting similar algorithms for the standard proximal gradient procedure.

4 - DEFBAL, Part II: New Approximate Augmented Lagrangian Methods and their Application to ADMM-Like Algorithms

Chang Yu, Rutgers University, Piscataway, NJ, United States, Jonathan Eckstein

We describe a new approximate augmented Lagrangian method in which the length of the multiplier adjustment step is dynamically adapted to the accuracy of the subproblem solution. In the DEFBAL framework, we combine this outer loop with an inner loop applying the FISTA-like method of Chambolle and Dossal to obtain an ADMM-like algorithm for convex programming. We also present some basic computational results.

SD31

Summit - 422

Discrete Optimization for Machine Learning

Invited Session

OPT: Machine Learning

Chair: Kayhan Behdin, LinkedIn, Sunnyvale

1 - Exact Algorithms for Regularized a-Optimal Experimental Design

Yongchun Li, Georgia Institute of Technology, Atlanta, GA, United States

We study the regularized A-optimal Experimental Design (AED) problem, which aims to select new data points, given an existing Fisher information matrix, to minimize the trace of the inverse of the overall Fisher information matrix. AED arises in various machine learning and statistics problems. We prove that the AED problem is NP-hard. To solve the AED problem effectively, we derive an equivalent mixed-integer second-order conic reformulation that can be directly solve by commercial solvers. Furthermore, we derive an equivalent mixed-integer convex reformulation that allows us to design a branch-and-cut algorithm. We provide theoretical comparisons of the continuous relaxations of the two exact formulations. Finally, we test our algorithms using real-world data to demonstrate their efficiency.

2 - Osscar: One-Shot Structured Pruning in Vision and Language Models with Combinatorial Optimization

Xiang Meng, MIT, Cambridge, MA, United States, Shibal Ibrahim, Kayhan Behdin, Rahul Mazumder

Structured pruning is a promising approach for reducing the inference costs of large vision and language models. By removing carefully chosen structures, e.g., neurons or attention heads, the improvements from this approach can be realized on standard deep learning hardware. In this work, we focus on structured pruning in the one-shot (post-training) setting, which does not require model retraining after pruning. We propose a novel combinatorial optimization framework for this problem, based on a layer-wise reconstruction objective and a careful reformulation that allows for scalable optimization. Moreover, we design a new local combinatorial optimization algorithm, which exploits low-rank updates for efficient local search. Our framework is time and memory-efficient and considerably improves upon state-of-the-art one-shot methods on vision models (e.g., ResNet50, MobileNet) and language models (e.g., OPT-1.3B -- OPT-30B). For language models, e.g., OPT-2.7B, OSSCAR can lead to 125x lower test perplexity on WikiText with 2x inference time speedup in comparison to the state-of-the-art ZipLM approach. Our framework is also 6x--8x faster. Notably, our work considers models with tens of billions of parameters, which is up to 100x larger than what has been previously considered in the structured pruning literature.

3 - Grand-Slamin&Rsquo; Interpretable Additive Modeling with Structural Constraints

Gabriel Afriat, Massachusetts Institute of Technology, Cambridge, MA, United States, Shibal Ibrahim, Kayhan Behdin, Rahul Mazumder

Generalized Additive Models (GAMs) are a family of flexible and interpretable models with old roots in statistics. GAMs are often used with pairwise interactions to improve model accuracy while still retaining flexibility and interpretability but lead to computational challenges as we are dealing with order of p^2 terms. It is desirable to restrict the number of components (i.e., encourage sparsity) for easier interpretability, and better computational and statistical properties. Earlier approaches, considering sparse pairwise interactions, have limited scalability, especially when imposing additional structural interpretability constraints. We propose a flexible GRAND-SLAMIN framework that can learn GAMs with interactions under sparsity and additional structural constraints in a differentiable end-to-end fashion. We customize first-order gradient-based optimization to perform sparse backpropagation to exploit sparsity in additive effects for any differentiable loss function in a GPU-compatible manner. Additionally, we establish novel non-asymptotic prediction bounds for our estimators with tree-based shape functions. Numerical experiments on real-world datasets show that our toolkit performs favorably in terms of performance, variable selection and scalability when compared with popular toolkits to fit GAMs with interactions. Our work expands the landscape of interpretable modeling while maintaining prediction accuracy competitive with non-interpretable black-box models. Our code is available at https://github.com/mazumder-lab/grandslamin.

4 - Modeling with Categorical Features via Exact Fusion and Sparsity Regularization

Kayhan Behdin, Massachusetts Institute of Technology, Cambridge, MA, United States

We study the high-dimensional linear regression problem with categorical predictors that have many levels. We propose a new estimation approach, which performs model compression via two mechanisms by simultaneously encouraging (a) clustering of the regression coefficients to collapse some of the categorical levels together; and (b) sparsity of the regression coefficients. We present novel mixed integer programming formulations for our estimator, and develop a custom row generation procedure to speed-up the exact off-the-shelf solvers. We also propose a fast approximate algorithm for our method that obtains high-quality feasible solutions via block coordinate descent. As the main building block of our algorithm, we develop an exact solver for the univariate case based on dynamic programming, which can be of independent interest. We establish new theoretical guarantees for both the prediction and the cluster recovery performance of our estimator. Our numerical experiments on synthetic and real datasets demonstrate that our proposed estimator tends to outperform the state-of-the-art.

5 - Mathematical Optimization Models for Multiple Support Vector Machine based classifiers

Harshit Kothari, University of Wisconsin-Madison, Madison, WI, United States, Victor Blanco, Jim Luedtke

We introduce a new approach based on Support Vector Machines (SVM) for building multiclass classifiers by arrangement of hyperplanes. We propose a mixed integer programming formulation for the problem. Additionally, we introduce symmetry breaking inequalities to strengthen the model. We also develop some dimensionality reduction methods which help us run the problem for larger datasets. Finally we present our results on an extensive set of computational experiments.

SD32

Summit - 423

Exact Approaches for Multistage Optimization

Invited Session

OPT: Global Optimization

Chair: Leonardo Lozano, University of Cincinnati, Cincinnati, OH, United States

1 - Decision Diagram Based Approaches for Discrete Bilevel Programming

Leonardo Lozano, University of Cincinnati, Cincinnati, OH, United States, Andre Cire, David Bergman

Integer bilevel programming problems are known to be very challenging due to the lack of strong relaxations that can be efficiently computed. We propose single-level representations of integer bilevel programming problems that rely on network flow-based approximations of the follower's value function, using decision diagram. We then show how we can derive scalable relaxations from this representation by constructing a minorizer of the follower’s value function. We experimentally compare our approach with state-of-the-art bilevel programming solvers and show that we can obtain competitive results for certain problem classes.

2 - A Single-Level Reformulation of Integer Bilevel Programs using Decision Diagrams

Sebastian Vasquez, Carnegie Mellon University, Pittsburgh, PA, United States, Leonardo Lozano, Willem-Jan van Hoeve

Integer bilevel programming problems are known to be very challenging due to the lack of strong relaxations that can be efficiently computed. We propose a new single-level representation of integer bilevel programming problems that relies on a network flow-based formulation of the follower’s value function, using decision diagrams and primal-dual constraints. We then show how we can derive scalable relaxations from this representation by incorporating only a restricted set of the master problem solutions, yielding dual bounds. We experimentally compare our approach with state-of-the-art bilevel programming solvers and show that we can obtain competitive results for certain problem classes, especially when considering follower's problems with combinatorial structure.

3 - Aggregation of two bilinear bipartite equality constraints and application to FEM update problem

Dahye Han, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey, Yang Wang

Aggregating constraints that describe a set gives a relaxation of the set. Some previous works on quadratic inequalities show that sometimes “nice” aggregations give a convex hull of the set. In this work, we wish to understand the aggregation of two bilinear bipartite equality constraints and whether we can find such nice aggregations. First, we show that for a special case where there are two variables involved, the convex hull can be represented as an intersection of at most three aggregations of constraints. Next, we show that when there are three variables involved, in general, an intersection of a finite number of aggregations cannot represent a convex hull. Furthermore, we provide a counterexample where an intersection of infinite aggregations still does not give a convex hull. Despite these results, aggregated equalities can provide a tight approximation of a convex hull. We experimentally show that adding a few aggregated equality constraints significantly reduces the size of the outer approximation of the convex hull of two bilinear bipartite equalities. Since adding many randomly aggregated constraints increases the complexity, we devise a simple heuristic to find aggregation weights based on a relaxed solution. We apply this heuristic to instances from finite element model updating problems in structural engineering. The experiment shows that the aggregated constraints can close almost 5% more gap in the root node and subsequently outperform the existing branch and bound method without aggregation and a commercial global solver.

SD33

Summit - 424

Advances in Stochastic Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Yifan Hu, EPFL & ETH, Zurich, Switzerland

1 - Contextual Stochastic Bilevel Optimization and Three-Stage Stochastic Programming

Yifan Hu, EPFL, Lausanne, Switzerland, Daniel Kuhn, Andreas Krause, Jie Wang, Yao Xie

Multistage stochastic programming poses significant challenges due to the curse of dimensionality. In this talk, we study three-stage stochastic programming using contextual stochastic bilevel optimization. Note that three-stage problems are often perceived as more challenging than the classical stochastic optimization. We develop a randomized gradient-based algorithm that achieves the same complexity bound as classical stochastic optimization. This result challenges conventional wisdom regarding the difficulty of three-stage problems, suggesting that prevailing beliefs may not always hold true. The development of such a randomized gradient-based algorithm opens up exciting directions for advancing algorithmic design for three-stage stochastic programming.

2 - Dynamic Pricing of Unique Products: Approximations and Applications to Bundling

Maxime Bouscary, Massachusetts Institute of Technology, Cambridge, MA, United States, Mazen Danaf, Saurabh Amin

Digital brokerage platforms present significant potential for improved load recommendations through strategic bundling and pricing strategies. By devising effective strategies factoring in carriers' strategic decisions and spatial positioning, these platforms can better accommodate carriers' preferences and reduce operational costs while ensuring timely load delivery. We formulate the problem as one of expected revenue maximization, considering a monopolistic online seller offering a set of unique and non-replenishable products priced dynamically over a finite horizon.

Our approach develops tractable bounds on expected revenue that are tailored to unique products and compatible with customer choice models that satisfy mild properties. Unlike the bounds provided by the static and fluid approximations, our bounds are asymptotically optimal under the multinomial logit model as the season length grows, yielding a performance bound that is $O(1/T)$. Subsequently, we propose a bundling algorithm leveraging these bounds to efficiently select bundles in a column-generation fashion. We analyze the performance of the developed bounds and the bundling algorithm on both synthetic and real-world data, allowing us to benchmark our bounds against the static and fluid approximations, and gain insight into the added value of bundling.

Our general formulation allows our framework to be used for a variety of applications. We show in our real-world study that the freight transportation industry can benefit from improved load recommendations and pricing decisions: by reducing not only operating costs, but also empty miles, our methodology benefits digital brokerage platforms while contributing to the industry's sustainability efforts.

3 - A Unified Linear Programming Framework for Reward Learning with Offline Human Demonstration and Feedback Data

Kihyun Kim, Massachusetts Institute of Technology, Cambridge, MA, United States, Jiawei Zhang, Pablo Parrilo, Asuman Ozdaglar

Inverse Reinforcement Learning (IRL) and Reinforcement Learning from Human Feedback (RLHF) are pivotal methodologies in reward learning, which involve inferring and shaping the underlying reward function of sequential decision-making problems based on observed human demonstration and feedback. Most prior work in reward learning has relied on prior knowledge or assumptions about decision or preference models, potentially leading to robustness issues. We introduce a novel linear programming (LP) framework tailored for offline reward learning. This framework estimates a feasible reward set from the primal-dual optimality conditions of a suitably designed LP, utilizing pre-collected trajectories without online exploration, and offers an optimality guarantee with provable sample efficiency. Our LP framework also enables aligning the reward functions with human feedback, such as pairwise trajectory comparison data, while maintaining computational tractability and sample efficiency. We demonstrate that our framework potentially achieves better performance compared to the conventional maximum likelihood estimation (MLE) approach through analytical examples and numerical experiments.

4 - An MILP-Based Solution Scheme for Factored and Robust Factored Markov Decision Processes

Man-Chung Yue, The University of Hong Kong, Hong Kong, Hong Kong, Huikang Liu, Wolfram Wiesemann

Factored Markov decision processes (MDPs) are a prominent paradigm within the artificial intelligence community for modeling and solving large-scale MDPs whose rewards and dynamics decompose into smaller, loosely interacting components. Through the use of dynamic Bayesian networks and context-specific independence, factored MDPs can achieve an exponential reduction in the state space of an MDP and thus scale to problem sizes that are beyond the reach of classical MDP algorithms. However, factored MDPs are typically solved using custom-designed algorithms that can require meticulous implementations and considerable fine-tuning. In this paper, we propose a mathematical programming approach to solving factored MDPs. In contrast to existing solution schemes, our approach leverages off-the-shelf solvers, which allows for a streamlined implementation and maintenance; it effectively capitalizes on the factored structure present in both state and action spaces; and it readily extends to the largely unexplored class of robust factored MDPs, whose transition kernels are only known to reside in a pre-specified ambiguity set. Our numerical experiments demonstrate the potential of our approach.

SD34

Summit - 425

Methods for Large-Scale Nonlinear and Stochastic Optimization II

Invited Session

OPT: Nonlinear Optimization

Chair: Jiahao Shi, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Shagun Gupta, UT Austin, Austin

Co-Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Newton-Cg Methods Under HÖLder Conditions

Chuan He, University of Minnesota, Minneapolis, MN, United States, Zhaosong Lu

Second-order methods have played an important role in the recent development of nonconvex optimization with numerous fruitful applications in science and engineering. In this talk, we introduce new Newton-CG methods for nonconvex unconstrained optimization with complexity guarantees under Hölder conditions. In particular, we propose the first parameter-free second-order method with iteration and operation complexity bounds that can seek an approximate first-order stationary point and second-order stationary point without requiring any prior knowledge of the Hölder parameters. Numerical results demonstrate the superior practical performance of our proposed method.

2 - Scalable Projection-Free Optimization Methods via Multiradial Duality Theory

Thabo Samakhoana, Johns Hopkins University, Baltimore, MD, United States, Benjamin Grimmer

Recent works have developed new projection-free first-order methods based on utilizing linesearches and normal vector computations to maintain feasibility. These oracles can be cheaper than orthogonal projection or linear optimization subroutines but have the drawback of requiring a known strictly feasible point to do these linesearches with respect to. In this work, we develop new theory and algorithms which can operate using these cheaper linesearches while only requiring knowledge of points strictly satisfying each constraint separately. Convergence theory for several resulting ``multiradial" gradient methods is established. We also provide preliminary numerics showing performance is essentially independent of how one selects the reference points for synthetic quadratically constrained quadratic programs

3 - A Simple, Reliable and Adaptive Trust-Region Method

Oliver Hinder, University of Pittsburgh, Pittsburgh, PA, United States, Fadi Hamad

We develop a simple trust-region method that obtains the best possible iteration bound for a second-order method on functions with Lipschitz continuous Hessian (up to additive logarithmic factors). Moreover, on the CUTEst test set our method is significantly faster and more reliable than the state of the art adaptive cubic regularization and trust-region methods implementations available in the GALHAD library.

4 - Fast regularized interior method for large scale separable convex quadratic programming

Ya-Chi Chu, Stanford University, Stanford, CA, United States

In this talk, we will present a new algorithm for convex separable quadratic programming (QP) called Nys-IP-PMM, a regularized interior-point solver that uses low-rank structure to accelerate solution of the Newton system. The algorithm combines the interior point proximal method of multipliers (IP-PMM) with the randomized Nyström preconditioned conjugate gradient method as the inner linear system solver. Our algorithm is matrix-free: it accesses the input matrices solely through matrix-vector products, as opposed to methods involving matrix factorization. It works particularly well for separable QP instances with dense constraint matrices. We establish convergence of Nys-IP-PMM. Numerical experiments demonstrate its superior performance in terms of wallclock time compared to previous matrix-free IPM-based approaches.

SD35

Summit - 427

Data-driven Decision-making in Urban Mobility

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Julia Yan, University of British Columbia, Vancouver, Canada

1 - Microtransit Design: Fixed-Line Transit, on-Demand Mobility, OR Both?

Shriya Karam, Massachusetts Institute of Technology, Cambridge, MA, United States, Alexandre Jacquillat, Julia Yan, Arthur Delarue

Urban mobility providers are increasingly leveraging emerging operating models based on microtransit, which integrate fixed-line transit and on-demand services. This paper seeks to understand the benefits and tradeoffs behind transit and on-demand operations and the optimal design of integrated microtransit systems. Using continuous approximation techniques, we first consider stylized settings to provide theoretical insights on the optimal transit and on-demand allocation structure. To build on our theoretical findings, we develop optimization models and algorithms to design spatial and temporal operations for door-to-door ride-sharing and fixed-line transit to serve passenger demand. Computational results show the advantages of operating hybrid transit and on-demand systems and the trade-offs across modes on agency operating costs and passenger convenience.

2 - Optimizing Public Transit: A Data-Driven Approach

Boxuan Zhou, UIUC, Champaign, IL, United States, John Birge, Ozan Candogan

Our model proposes an enhancement for fixed-station public transportation systems by optimizing express station selection and vehicle allocation between express and regular routes. While applicable to both bus and subway systems, we focus on subways due to data availability. The problem of optimizing express routes is inherently complex and algorithmically challenging. To simplify it, we introduce an "admissibility" condition that narrows down the feasible express route set to a practically relevant class. Leveraging this condition, we develop a dynamic programming algorithm for a manageable solution to the planner's problem. Initially designed for a single straight train line, our algorithm extends successfully to various transportation network topologies, including circular routes and tree structures.

Our methodology requires knowledge of riders' origin-destination (O-D) demand, which is often absent in subway systems. What we have is turnstile data that captures station entrances and exits. To bridge this gap, we propose a choice model that accounts for riders' destination preferences, time, and distance values. Algorithms are provided to estimate this model and construct the O-D demand matrix.

We apply our approach to the NYC subway system using data from the Metropolitan Transportation Authority (MTA) and our estimation framework for O-D demand. Through dynamic programming, we determine optimal express station locations and frequencies. Focusing on the Red (Green) Line, our results indicate potential time savings for riders ranging from 5% to 15% (2% to 5%), and the optimal express station locations differ significantly from the current setup.

3 - Reweighted L1-Norm Minimization Approach to Transportation Network Design Under Demand Endogeneity

Pushpendra Singh, Dartmouth College, Hanover, NH, United States, Vikrant Vaze

Many transportation applications involve some form of network design. Network design problem has been well-studied in the literature but most prior studies assume passenger demand to be exogenously determined. In reality, demand depends on network design. But capturing such endogenous demand adds significantly to the model complexity. We define a new class of mixed-integer non-convex optimization problems for network design, which are important to practice across many fields and share some mathematical similarities that motivate this work. We provide a compact and general formulation for this class of problems and also define subproblems and map existing research to these subclasses. Next, we characterize the key existing solution approaches, as well as presenting new approaches, for solving these problems successfully. The approaches include: Reweighted L1 norm minimization, Exponential Conic Programming, and Adaptive Discretization Algorithm. While all the three approaches provide either optimal or very close to optimal solution, the reweighted L1-norm minimization which takes the advantage of the sparse nature of matrix of transportation network design problems, outperforms the other approaches when it comes to scalability while providing very close to optimal solutions (within 1.5% optimality gap). Our results from both real-life case studies and simulated networks uphold our claims while providing solutions in much faster runtimes than other proposed methods.

4 - Achieving Transit Accessibility Gains through Flexible Bus Routing and Advance Online Reservation for a Low-Demand Route

Md Hishamur Rahman, University of Utah, Salt Lake City, UT, United States, Ye Chen, Shijie Chen, Yanshuo Sun, Nikola Markovic

The objective of this paper is to enhance the accessibility of conventional bus routes by increasing the number of stops, particularly on routes where existing stops are underutilized. We propose a new type of semi-flexible transit system where buses can selectively skip stops lacking passenger demand and take shortcuts, thereby reducing the tour duration compared to that of conventional routes. This system requires passengers to provide advance notice for their desired stops, ensuring these stops are reliably serviced. We introduce analytical and simulation-based models to design and optimize this semi-flexible transit system. The optimization objective is to maximize the number of stops, subject to tour duration and arrival time constraints. Integral to our methodology is the development of an efficient online learning approach, which we use to sequentially evaluate the system's feasibility across different scenarios. In the Allegany County, Maryland case study, significant enhancements were observed in bus routes where route configurations allowed taking substantial shortcuts and the average probability of skipping a stop exceeded 45%. In these instances, the number of stops can be increased by up to 160%, with the actual improvement depending on route configuration, passenger demand, and advance notice requirements.

5 - Curbing Emissions: Enhancing Sustainability Through Collaborative Shipment in Horizontal Supply Chains

Tanvir Ibna Kaisar, University of Southern California, Los Angeles, CA, United States, Maged Dessouky, Randolph Hall

In recent years, the importance of sustainable and resilient supply chains has become increasingly important due to the growing concerns over energy crises, environmental pollution, and supply chain disruptions. Amidst these challenges, battery electric vehicles (BEVs) have emerged as a promising solution as they offer several key advantages that contribute to sustainable and resilient supply chains. BEVs have the potential to significantly reduce greenhouse gas (GHG) emissions in the logistics sector, thereby helping companies meet their sustainability goals and mitigate their environmental impact. Moreover, BEVs can enhance supply chain resiliency by reducing reliance on volatile fossil fuel markets and providing a more stable and predictable energy source for transportation. However, the BEV technology is still evolving, and the significant initial investment required for BEV adoption remains a deterrent, particularly among small companies. To address this challenge, we propose a cooperative mechanism for BEV adoption among multiple small companies in a horizontal supply chain. We specifically focus on the logistics sector, where a set of firms deliver their products to their customers through a shared distribution center using a BEV fleet. We formulate the routing problem as a mixed integer programming (MIP) problem and solve it exactly for small instances. For larger instances, we use a customized Adaptive Large Neighborhood Search (ALNS) heuristic. We also develop cost-sharing mechanisms to fairly distribute the cost of the collaboration among the participants under various scenarios.

SD36

Summit - 428

Optimization and Decarbonization of Multi-modal Freight Transportation Systems

Invited Session

TSL: Freight Transportation

Chair: Adrian Hernandez, Northwestern University, Evanston, IL, United States

1 - Multimodal Freight Network Design and Facility Location Optimization to Support Freight Transportation Decarbonization

Adrian Hernandez, Northwestern University, Evanston, IL, United States, Pablo Durango-Cohen, Hani Mahmassani, Vasileios Volakakis, Cal Skiles, Dana Monzer

This paper proposes a multimodal freight network design and facility location model to optimize the integration of road, rail, and maritime transportation modes with respect to efficiency, costs, and emissions. We build on classical network models such as multi-commodity network flow problems and hub location-allocation problems to formulate a model for (i) locating transloading facilities and (ii) establishing links to connect two or more freight modes and networks. We present model variants with different objectives, including minimization of transportation costs, minimization of environmental impacts, and maximization of connectivity across modes. Furthermore, we capture alternative refueling facility location decisions within the model and allowing flows to be split among different modes and energy technologies between nodes on the network. We apply this model in the context of the US freight transportation network and market, with examples including the deployment of charging infrastructure to support battery-electric trucks and locomotives, as well as the location of transloading facilities to connect rail and domestic maritime modes. In addition to evaluating model results on cost, emissions, and travel time impacts, we evaluate network resilience through simulation of local and regional disruptions.

2 - Improving Multimodal Freight Transportation Operations: A Hybrid Multi-Objective Routing Optimization – Agent Based Modelling Logistics Framework

Vasileios Volakakis, Northwestern University, Evanston, IL, United States, Adrian Hernandez, Max Ng, Cal Skiles, Dana Monzer, Pablo Durango-Cohen, Hani Mahmassani

This work proposes a logistics optimization framework for multimodal freight operations, consisting of two methodological components, an optimization-based and a simulation-based model. Regarding the former component, a multi-objective routing optimization model with objectives of travel time, emissions, and risk (disruption and charging/refueling range sufficiency – related) minimization was developed, that provides a set of pareto optimal solutions, i.e. multiple routing options. For the solution of the optimization problem, the Adaptive Weighted Sum (AWS) method has been used, which initially uses the common weighted sum approach to swiftly approximate the Pareto surface and identify a grid of Pareto front segments. These segments are then refined through additional equality constraints that link specific reference points to anticipated optimal solutions on a segmented hypersurface in a multidimensional objective space. For the later methodological component, an agent-based model was developed, to simulate the transport operations and capture behavioral responses of transport agents, focusing on key variables in the mode choice process, such as price, transit time, reliability, and flexibility. Such agents are shipper, carrier, and container agents. The proposed methodological framework is to be applied at a national level set of networks and scenarios, with the ultimate objective of forging a tool that facilitates the comparison and evaluation of various routing options under different scenarios, accounting for the interrelations and interactions of different entities in such a network.

3 - A Decarbonized and Resilient Intermodal Freight Transportation (DRIFT) Modeling Platform for Intermodal Logistical Decisions under Uncertainty

Kuilin Zhang, Michigan Technological University, Houghton, MI, United States, Yun-Chu Hung, Yingtong Tan

This talk presents an intermodal logistical modeling platform - Decarbonized and Resilient Intermodal Freight Transportation (DRIFT) to support long-term planning decisions on mode, route, and charging for future charging/fueling strategies up to the year 2050, and also quasi-real-time dynamic decisions on scheduling and re-routing to respond to disruptions and real-time traffic conditions. The objective is to reduce overall energy usage per ton-mile of freight transported while minimizing the levelized cost of ton-kilometers (LCOTKM) with increased supply chain resiliency. Considering future scenarios of renewable energy sources, charging/fueling infrastructure rollout strategies, projected freight demand from ports to warehouses, and natural- and man-made extreme events for intermodal logistics, DRIFT provides optimized intermodal logistical decisions that could help identify the most cost-effective transition pathways to a net-zero GHG emissions intermodal transportation systems, including water, rail, and road.

4 - Identifying capacities in a multimodal maritime freight network

Kyle Bathgate, The University of Texas at Austin, Austin, TX, United States, Debojjal Bagchi, Stephen Boyles

We investigate how a "capacity" can meaningfully be defined in multimodal maritime freight systems, and how such a notion can be used to improve performance and resilience by identifying bottlenecks. Shipping channels and ports are each complex systems that interact deeply, and the bottlenecks in the individual facilities may not represent bottlenecks in the combined multimodal system. We first describe a data-driven discrete event simulation framework to describe operations of such a system. We then show how this simulation framework can provide a notion of "capacity" by identifying the most significant variables and parameters, and how this notion can be used in resource allocation problems involving maintenance or improvements in multimodal systems. We present a case study to demonstrate how the framework can be applied.

5 - Integrated multimodal network models for decarbonization of intermodal freight shipments

Yiduo Huang, University of California-Berkeley, Berkeley, CA, United States, Yanfeng Ouyang

The U.S. freight transportation industry is responsible for a significant portion of the nation’s greenhouse gas (GHG) emissions, and its under-capacitated infrastructure has exacerbated its vulnerability to disruptions. We present integrated data and computational models that enable modeling and decision making on the nation’s intermodal freight logistics systems across water, roadway, and railway modes. Logistics optimization models will be developed that could explain freight shippers’ choices (e.g., mode, origin/destination) and logistics service providers’ decisions (e.g., on routing and scheduling) across multiple integrated shipment modes.

SD37

Summit - 429

Game Theory in Smart Mobility

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Xiaotong Sun, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, N/A

Co-Chair: Jiayang Li, Northwestern University, Evanston, IL, United States

1 - Can information design resolve EV charging chaos? Author: Qianni Wang, Jiayang Li, Xiaotong Sun, Yu (Marco) Nie

Qianni Wang, Northwestern University, Evanston, IL, United States

The rapid expansion of the electric vehicle (EV) market has heightened concern about the shortage of charging infrastructure, especially for long-distance travel. Range anxiety, combined with uncertain availability and delays at charging facilities, can lead to behaviors that undermine system efficiency, at times when efficiency matters the most. Here, we consider a designer who, equipped with exclusive real-time information on the uncertain state of charging, attempts to guide EV drivers toward a socially desirable outcome by designing an optimal information dissemination scheme. Building on information design theory, we develop Bayesian persuasion models for both public and private information designs. Whereas the public model provides the same information to all drivers, the private model offers personalized recommendations based on individual attributes (i.e. the remaining range of an EV at decision points in our model). Our analysis reveals that consistently disclosing the true state of charging is the most effective scheme for public information design. However, personalized information delivery often leads to better outcomes. We also find that information design works best when there is a high level of uncertainty in the drivers' perception of the charging state, and that benefits of information tend to increase as the cost associated with charging worsens relative to other choices.

2 - Sequential Information Design: Markov Persuasion Process and Its Efficient Reinforcement Learning

Jibang Wu, University of Chicago, Chicago, IL, United States

This paper proposes a novel model of sequential information design, namely the Markov persuasion processes (MPPs), in which a sender, with informational advantage, seeks to persuade a stream of myopic receivers to take actions that maximize the sender's cumulative utilities in a finite horizon multi-agent Markovian environment. Planning in MPPs thus faces the unique challenge in finding a signaling policy that is persuasive to the myopic receivers and simultaneously induces the optimal long-term cumulative utility for the sender. For such a problem, we design a provably efficient no-regret learning algorithm, which features a novel combination of both optimism and pessimism principles. In particular, we obtain optimistic estimates of the value functions to encourage exploration under the unknown environment, and additionally robustify the signaling policy with respect to the uncertainty of prior estimation to prevent receiver's detrimental equilibrium behavior. Our algorithm enjoys sample efficiency by achieving a sublinear $\sqrt{T}$-regret upper bound. Furthermore, both our algorithm and theory can be applied to MPPs with a large space of outcomes and states via function approximation, and we showcase such a success under the linear setting.

3 - Balancing mobility and power grid considerations with shared autonomous electric vehicles

Stephen Boyles, The University of Texas at Austin, Austin, TX, United States

Authors: Jake Robbennolt, Meiyi Li, Stephen D. Boyles, Javad Mohammadi Abstract: Shared autonomous electric vehicle systems can provide on-demand mobility, but also interact extensively with the power grid. To balance these needs, we describe a rolling-horizon model-predictive control approach to vehicle dispatch, accounting for both transportation network dynamics and power distribution constraints. We then present a distributed solution technique using the alternating direction method of multipliers to improve computational efficiency and support real-time implementation. The transportation and power flow subproblems are coupled by power transfer at charging stations, and the transportation subproblem is further decomposed across vehicles. In this cooperative game scenario, we investigate incentive compatibility of the integrated mobility-energy system, computational performance of our algorithms, and impacts on mobility and grid operation through numerical experiments on urban test cases.

4 - Safety, Liability, and Insurance in the Age of Automated Driving

Sina Bahrami, University of Michigan, Ann Arbor, MI, United States, Daniel Vignon

This paper investigates two questions related to safety and insurance in the age of automation. First, the question of safety and liability under infrastructure-assisted automated driving. In such an environment, automakers provide vehicle automation technology while infrastructure service providers (ISSPs) provide smart infrastructure services. Additionally, customers can receive coverage for accidents from either of these actors but also from legacy auto insurers. We investigate the effect of market structure on safety and accident coverage and show that an integrated monopoly provides full coverage and fully accounts for accident costs when choosing safety levels. However, in the Nash setting, even though full coverage obtains, lack of coordination leads to partial internalization of accident costs by the automaker. Moreover, multiple equilibria might exist, some of them undesirable. We show that, with and without legacy insurance, an appropriate liability rule can induce optimal safety levels under the Nash setting. Our second question concerns itself with the role of legacy auto insurance in the age of infrastructure-assisted automated driving. We show that the industry is not necessary for optimal coverage when the cost of accidents is known in advance and all possible accident scenarios are contractible. In fact, their presence can even harm safety, even though it ensures full coverage for accidents. However, when only insurance contracts with capped liability for automakers and ISSPs are available, we cannot rule out that customers benefit from legacy insurers. Thus, the disappearance of the industry in the age of automated driving is not a foregone conclusion.

5 - Error bounds for stochastic user equilibrium traffic assignment

Debojjal Bagchi, The University of Texas at Austin, Austin, TX, United States, Stephen Boyles

Author: Debojjal Bagchi, Stephen Boyles Abstract: In stochastic user equilibrium traffic assignment, we develop bounds on the distance between a given feasible solution and the equilibrium solution in terms of a gap function. The intent is to provide guidance on termination criteria to reduce run times, which is important because this assignment problem is often a subproblem to a more complex bilevel optimization. These mathematical bounds complement existing rules-of-thumb drawn empirically from numerical case studies. Our approach is based on Taylor's theorem, as applied to the fixed-point formulation of the stochastic user equilibrium assignment, and provides upper bounds on differences in both aggregate metrics (total travel time, distance traveled) and disaggregate metrics (link flows, path flows). We demonstrate that these bounds are tight and cannot be further improved without additional restrictions on the network topology or problem instance.

SD38

Summit - 430

AAS Student Presentation Competition II

Award Session

Aviation Applications

Chair: Wayne Ng, Singapore University of Technology and Design, Singapore, Singapore

1 - Energy-constrained, Risk-averse Drone Delivery under Spatiotemporally Varying Winds

Minghao Chen, Columbia University, New York, United States, MAX LI, Marco Giometto, Andrew Smyth

2 - Dynamic Discretization Discovery Algorithm for Airline Timetable Development and Fleet Assignment with Passenger Choice

Ritesh Ojha, Georgia Institute of Technology, Atlanta, GA, United States, Luke Marshall, Keji Wei

The airline timetable development and fleet assignment problem un- der endogenous passenger choice (ATFP), aims to create flight timetables based on passenger choice while optimizing fleet utilization for profitability [1]. Traditionally, airlines make incremental adjustments to their existing flight schedules. State-of-the-art methods for ATFP involve multiphase heuristic ideas. In this paper, we develop an exact dynamic discretization discovery (DDD) algorithm that iteratively solves relatively small mixed-integer programs (MIPs) formulated over partially time-expanded networks to generate optimal solutions [2]; the effectiveness of the algorithm relies on a sophisticated lower-bound MIP model and a novel upper-bound construction procedure. We prove that the DDD algorithm always converges to the optimal solution in a finite number of iterations. We also develop effective arc-based and flight itinerary-based network refinement strategies that enable the DDD algorithm to converge to the optimal solutions while generating only a small fraction of the complete time-expanded network.

The exact DDD algorithm finds the optimal solution and is 8× faster than directly using a commercial solver on small-scale instances obtained from real-life data of Alaska Airlines; the commercial solver cannot prove optimality for the small instances even in 15 minutes. The algorithm is computationally efficient because it solves small MIPs (LBM) in each iteration and finds the optimal solution while generating only 10−30% of the complete time-expanded network. Computational results suggest that our solution approach can result in significant profit improvements, as compared to the most advanced incremental and multi- phase heuristic approaches to flight timetabling.

3 - Two-stage Drone Delivery Network Design with Uncertain Demand

4 - Optimization-Based Simulation of Target Time Over For Flights Within the Decentralized ASEAN Region

SD39

Summit - 431

Railway Scheduling and Management of Crews, Intermodal, Yards and Networks

Contributed Session

Railway Applications

Chair: Hangyu Ji, N/A

1 - Benders Decomposition for Robust Tactical Railway Crew Scheduling

Bart van Rossum, Erasmus University Rotterdam, Rotterdam, Netherlands, Twan Dollevoet, Dennis Huisman

We consider robust tactical crew scheduling for a large passenger railway operator, whose timetable is subject to minor but uncertain updates on a weekly basis. Before the start of the planning period we must select a number of templates, specifying the time windows in which crew is supposed to work. These templates must be cost-efficient whilst providing sufficient capacity for the realised updated timetables. Moreover, the templates must be compatible with the rostering process, i.e., we wish to limit the number of different templates used and ensure that there are not too many templates on early or late moments of the day. Throughout the planning period, the templates are fixed and the updated timetable is released on a weekly basis. All tasks in the timetable must be covered using the capacity provided by the templates, or, if necessary, costly reserve capacity. We model the template selection problem as a scenario-based robust optimisation model, for which we propose a Benders decomposition algorithm. We conduct computational experiments on real-life instances from Netherlands Railways and show that our algorithm is able to solve instances featuring up to three scenarios and up to 1,000 tasks per scenario to optimality. Including more scenarios, allowing for more flexibility in choosing templates, or enlarging the template length all lead to reductions in reserve crew costs without increasing template-based crew costs.

2 - An Open-Specification-Based Ecosystem for Optimizing intermodal Freight Transportation

Xuesong (Simon) Zhou, ASU, Tempe, AZ, United States

Our presentation introduces the GMNS-based Open-Specification-Based Planning approach, aiming to achieve Net-Zero Emissions through sustainable supply chains. We will highlight the Open Computational Ecosystem for Adaptive Networked National and Transportation Systems (OCEAN+), a crucial tool that employs dynamic, multi-level structures and scalable computing to tackle complex scheduling and routing challenges in intermodal freight operations. This initiative promises significant energy savings and aims to set new standards for sustainability and resilience in the freight sector.

3 - Railcar-to-train Optimization in a Freight Railyard

Ruonan Zhao, Texas A&M University, College Station, TX, United States, Joseph Geunes, Xiaofeng Nie

A typical freight railyard at a manufacturing facility contains multiple tracks designated for classification and departure purposes. After loading cars with outbound goods at the manufacturing facility, railcars with different destinations are stored on these classification tracks and must later be moved to departure tracks to create outbound trains with designated destinations, incurring associated repositioning costs. Car storage locations on classification tracks are often determined before knowledge of destination locations is revealed, leading to problems that require retrieving subset of cars located randomly throughout the yard to build outbound trains. To address this challenge, we propose a mixed-integer programming (MIP) model that allows simultaneous movement of multiple railcars at minimum repositioning cost. By constructing the associated conflict graph, we prove the problem's NP-hardness. Furthermore, we devise a Track-Adaptive Car Reassignment Heuristic (TACR-H) algorithm to solve the problem. Results of a set of numerical experiments are presented to demonstrate the ability of the TACR-H algorithm to obtain fast solutions while achieving acceptable solution quality.

4 - Joint Planning of Passenger and Freight Trains in High-Speed Railway Express System

Hangyu Ji, Beijing Jiaotong University, Beijing, China, People's Republic of, Jiateng Yin

High-speed railway (HSR) has experienced significant expansion and has become the primary choice for passengers traveling between intercities in recent years. In HSR, passenger demand varies during different time periods. Thus, the rail managers, suffering from financial pressures in low seasons, are actively exploring the potentials for using a part of vehicles in some trains to transport both passengers and fast-delivery goods, while these is few research along this line. Our study formally addresses this problem for the integrated planning of train schedule and set allocations for passengers and freight. We develop a mixed integer linear programming model to optimally determine the tactical train schedule, train stopping plan, and train compositions. The objective of our model is to minimize the total travel time and operation costs of trains while satisfying the demands of both passengers and freight. In addition, we carefully consider some new constraints in our formulation, involving the time window limitation of freight, the freight loading/unloading work time at platforms, which are practically significant while were not fully investigated in the existing literature. Then, we prove that the constructed model is an NP-hard problem, indicating its inherent difficulty in being solved. To tackle this challenge, we develop an decomposition-based approach to solve the model, where the passenger-freight allocation is formulated as the master problem, and the scheduling of each train is formulated as one subproblem. Real-world case studies on Wuhan-Guangzhou high-speed railway corridor are conducted to verify the effectiveness of our approach.

5 - Train Service Design Problem with Perceived Headway Consideration and Passenger Flow Allocation for Urban Rail Transit Lines

Max Ng, Northwestern University Transportation Center, Evanston, IL, United States, Hani Mahmassani, Omer Verbas, Taner Cokyasar, Draco Tong

This study addresses the train service design problem for urban rail transit lines, optimizing stop sequences and headways of each pattern simultaneously to minimize user costs (composed of riding, waiting, and transfer times) under operational constraints (e.g., number of trains). The formulation does not rely on pre-defined candidate pattern sets but allows for flexible pattern options. It incorporates various service strategies (e.g., express/local services, short-turning, and dead-heading) and considers bi-directional origin-destination (inelastic) demand and (continuous) operations. It evaluates perceived headways of joint patterns for passengers, assigns passenger flows to each pattern accordingly, and allows transfers across patterns in different directions. Multi-line optimization of patterns and headways considering fleet assignment is also considered. We develop a destination-labeled multi-commodity network flow (MCNF) formulation, leading to efficient mixed-integer linear programming (MILP) solutions. Binary variables assign stop sequences and headways to patterns, and perceived headways to origin-destination pairs. Continuous variables assist in evaluating flows and costs. The methodology is demonstrated with a case study of metro lines in Chicago, USA, with 42 stations and scenarios ranging from two to three patterns and three to seven headways. Results indicate that optimized patterns outperform a full-stop pattern, reducing average user costs by 2.4% and average riding time by 9.0%.

6 - Fast and Efficient Heuristic Algorithm for Selecting Highway-Rail Grade Crossings for Closure

Razieh Khayamim, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Bokang Li, Payam Afkhami, Zeinab Elmi, Ren Moses, John Sobanjo, Eren Ozguven, Maxim Dulebenets

An intersection of railway and highway segments at the same elevation, known as a highway-rail grade crossing or level crossing, presents significant safety concerns and operational challenges. To mitigate these issues, different methods can be adopted, such as implementation of countermeasures (e.g., installation of flashing lights, gates, and median barriers), grade separations, and level crossing closures. The closure of selected level crossings, when done thoughtfully in consultation with relevant stakeholders and affected communities, can be a highly effective alternative. However, the majority of studies in this field primarily focuses on the implementation of countermeasures and mainly concentrates on safety-related aspects. Considering the drawbacks in the current research, a novel optimization model to identify level crossings for closure is presented in this study, which evaluates various benefits from closing specifically selected level crossings, including safety, economic, and environmental benefits, as well as reduction in traffic delays and maintenance and operational costs. A specialized heuristic algorithm inspired by sorting principles is developed to solve the decision problem addressed in this study. A set of comprehensive computational experiments are performed focusing on level crossings situated in the State of Florida (United States). The findings show clear superiority of the developed heuristic algorithm against the exact method, as it was able to solve the generated problem instances ≈20 times faster with the average optimality gap of only 0.14%. Additionally, various sensitivity analyses are performed to illustrate managerial implications from the implementation of the proposed optimization approach to administer closures of level crossings.

SD40

Summit - 432

Socially Responsible Behavioral OM

Invited Session

Behavioral Operations Management

Chair: Sebastian Villa, University of New Mexico, Albuquerque, NM, United States

Co-Chair: Vinit Tipnis, Kelley School of Business, Indiana University Bloomington, Bloomington, IN, United States

1 - Time Accommodation for Call Center Agents with Disabilities: A Mediation Analysis

Yingru Han, University of South Carolina, Columbia, SC, United States, Luv Sharma, Sriram Narayanan

While labor shortage remains persistent in the global economy, a subsegment of the population - people with disabilities (PWDs), has been continuously marginalized by the labor market. Positions exist where disabilities do not directly prevent individuals from performing the job, such as a call center agent. However, the high-stress nature of the job requires appropriate accommodations for PWDs to sustain and prosper, which are not often formally implemented especially in less developed economies. Such a phenomenon requires employers and PWDs to create informal accommodations. In this study we examine the role of spontaneous idleness, defined as the naturally occurring idleness between two calls, as a new source of accommodation for PWDs working in call centers. We first study how spontaneous idleness affects PWDs' productivity and how this relationship changes for PWDs of different levels of disability severity. We also study how spontaneous idleness affects agents' behaviors in taking flexible breaks, defined as the flexibility given by employers to take breaks as an agent sees needed. Using panel data at an agent-day level and a mediation analysis, we uncover the positive relationship between spontaneous idleness and productivity and identify flexible breaks as a partial mediator and the severity level of disability as a moderator. Our findings can assist organizations in similar contexts to better accommodate PWDs and improve their productivity by configuring spontaneous idleness.

2 - Assigning Volunteers to Beneficiaries in A Non-Profit Organization

Shikha Safaya, Scheller College of Business, Georgia Tech, Atlanta, GA, United States, Basak Kalkanci, Ravi Subramanian

Non-Profit Organizations (NPOs) play an important socioeconomic role by providing services to people in need, often with volunteers from the community acting as service providers. However, absent monetary incentives, volunteer scheduling and retention are key challenges faced by many NPOs. Motivated by the context of Meals on Wheels Atlanta (MOWA), an NPO providing meals and personal interaction through volunteers to elderly with limited mobility and food insecurity, we examine the incorporation of volunteer preferences for task assignment as a lever to induce provision of services and enhancement of service quality. The populations of volunteers [seniors] are of two types: individuals who prefer the base service of only delivering [only receiving] meals (type B), and individuals who are open to offering [who value] interactions alongside meal delivery (type I). We compare two policies: (i) a pooled policy, which groups volunteers together, irrespective of their task preferences (to represent the status quo); and, (ii) a preference-based policy, which incorporates volunteer preferences (as a potential improvement over the status quo). We explore the tradeoff between incorporating volunteer preferences during task assignment and pooling all volunteers to alleviate supply-demand mismatch. In doing so, we endogenize the volunteers' decisions to participate in the delivery activities at MOWA based on their expectations of the utility from serving as well as their outside options. We contrast the performance outcomes of alternative matching policies and suggest ways by which NPOs can better align incentives of all stakeholders.

3 - Enhancing Volunteer Retention: The Role of Experienced Volunteers

Vinit Tipnis, Kelley School of Business, Indiana University Bloomington, Bloomington, IN, United States, Christopher Chen, Fei Gao

Volunteers play a pivotal role in the daily operations of nonprofit organizations (NPOs), yet retaining volunteers remains a critical challenge for NPOs. This study explores a cost-effective approach to improving volunteer retention by leveraging the potential of the existing pool of experienced volunteers. Specifically, we examine how interactions between experienced and new volunteers impact the latter's retention behavior. Using data of 45,315 volunteers at a food bank over a five-year period, we estimate logistic regression model to explore interaction patterns among the volunteers. Our findings indicate that the presence of experienced volunteers in a volunteering shift increases new volunteer retention by 15.5%. Notably, this positive effect on retention is experienced by all new volunteers, regardless of whether they volunteered alone or as part of a group and across shifts with or without beneficiary interactions. However, through network analysis, we find that increased familiarity among experienced volunteers negates this positive effect. Our study contributes to the volunteer management literature by highlighting the importance of social interactions for volunteer retention and how NPOs can utilize their existing experienced volunteers to enhance new volunteer retention.

SD41

Summit - 433

Algorithm Design for Randomized Experiments

Invited Session

Applied Probability Society

Chair: Christopher Harshaw, MIT / UC Berkeley, Cambridge, MA, United States

1 - Causal Inference & Interference in Bipartite Experiments

Jean Pouget-Abadie, Google, New York, NY, United States

When the treatment assignment of one unit affects the outcome of another, we say there is interference. Interference is especially prevalent in marketplaces, where buyer and seller interactions lead to complex dependence structures. As a violation of the stable unit treatment value assumption, the presence of interference can lead to bias of standard estimators under naive randomized designs. In this talk, we will cover a set of design and estimation paradigms used at Google to conduct causal inference research in a bipartite graph setting, inspired from, but not limited to, marketplace experiments, with specific attention paid to clustered randomized designs.

2 - Estimating Direct Effects under Interference: a Spectral Approach to Experimental Design

Anthimos Vardis Kandiros, MIT, Cambridge, MA, United States, Constantinos Daskalakis, Christopher Harshaw, Charilaos Pipis

From clinical trials to corporate strategy, randomized experiments are a reliable methodological tool for estimating causal effects. In recent years, there has been a growing interest in causal inference under interference, where treatment given to one unit can affect outcomes of other units. While the literature on interference has focused primarily on unbiased and consistent estimation, designing randomized network experiments to ensure tight rates of convergence is relatively under-explored for many problems.

In this talk, we study the problem of direct effect estimation under interference. Here, the interference between experimental subjects is captured by a network and the experimenter seeks to estimate the difference between the outcomes when (i) a unit is treated and its neighbors receive control and (ii) the unit and its neighbors receive control. We propose a Horvitz—Thompson style estimator and a new experimental design under which the normalized variance is bounded as $n * Var <= O(\lambda)$, where λ is the largest eigenvalue of the adjacency matrix of the graph. This experimental design, which is inspired by techniques from spectral graph theory, establishes the best known rate of convergence for this problem; in fact, we offer lower bounds for any experimental design, which match our rates in certain instances. In addition, we present a variance estimator and CLT which facilitate the construction of asymptotically valid confidence intervals. Finally, simulations using data from a real network experiment corroborate the theoretical claims.

3 - Finite Population Regression Adjustment and Non-Asymptotic Guarantees for Treatment Effect Estimation

David Arbour, Adobe Research, Charlottesville, VA, United States

The design and analysis of randomized experiments is fundamental to many areas, from the physical and social sciences to industrial settings. Regression adjustment is a popular technique to reduce the variance of estimates obtained from experiments, by utilizing information contained in auxiliary covariates. While there is a large literature within the statistics community studying various approaches to regression adjustment and their asymptotic properties, little focus has been given to approaches in the finite population setting with non-asymptotic accuracy bounds. Further, prior work typically assumes that an entire population is exposed to an experiment, whereas practitioners often seek to minimize the number of subjects exposed to an experiment, for ethical and pragmatic reasons. In this work, we study the problems of estimating the sample mean, individual treatment effects, and average treatment effect with regression adjustment. We propose approaches that use techniques from randomized numerical linear algebra to sample a subset of the population on which to perform an experiment. We give non-asymptotic accuracy bounds for our methods and demonstrate that they compare favorably with prior approaches.

4 - Causal Inference Under Interference Through Designed Markets

Evan Munro, Stanford University, Stanford, CA, United States

In many markets a centralized mechanism determines the allocation of goods. When an individual-level intervention affects submissions to the mechanism, program evaluation is challenging due to spillover effects that occur through the mechanism. For example, a bidder-level treatment that increases some bids in an auction will impact the market-clearing price, affecting all auction participants. We show that if the mechanism is truthful and has a ``cutoff'' structure, then interference is constrained, and it is possible to estimate the Global Treatment Effect (GTE) under a selection-on-observables assumption. Our proposed estimator is doubly-robust and has an asymptotic variance that meets the semi-parametric efficiency bound. We also characterize heterogeneous treatment effects under interference in this setting and propose estimators for the optimal targeting rule. Taking into account equilibrium effects notably diminishes the estimated impact of an information intervention on inequality in the Chilean school system.

SD42

Summit - 434

Statistical Reinforcement Learning

Invited Session

Applied Probability Society

Chair: Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

Co-Chair: Raaz Dwivedi, Cornell Tech, Cornell ORIE, New York City, NY, United States

1 - Temporal Difference Versus Monte Carlo: An Optimal and Adaptive Resolution

Wenlong Mou, University of Toronto, Toronto, ON, Canada

Temporal difference (TD) and Monte Carlo (MC) are two basic strategies for value learning in Markov decision processes. Each approach offers unique advantages -- TD is believed to enjoy smaller asymptotic variance; while MC yields an unbiased estimator valid for any sample size. In reinforcement learning, a long-standing puzzle is to reconcile the merits of both principles, and develop algorithms with best-of-both-worlds guarantees.

In this talk, I will discuss some recent advances in resolving this puzzle. Focusing on statistical properties with a finite sample size, we develop a new class of algorithms that optimally interpolates between TD and MC, and adapts to the difficulties of problem instances. Among other results, I will highlight the impact of local structures of the Markov chains on the fundamental complexities of estimation.

2 - Settling the Sample Complexity of Online Reinforcement Learning

Zihan Zhang, Princeton University, Princeton, NJ, United States

A central issue lying at the heart of online reinforcement learning (RL) is data efficiency. While a number of recent works achieved asymptotically minimal regret in online RL, the optimality of these results is only guaranteed in a "large-sample" regime, imposing enormous burn-in cost in order for their algorithms to operate optimally. How to achieve minimax-optimal regret without incurring any burn-in cost has been an open problem in RL theory.

We settle this problem for finite-horizon inhomogeneous Markov decision processes.Specifically, we prove that a modified version of MVP (Monotonic Value Propagation), and achieves the minimax lower bound for the entire range of sample size, essentially eliminating any burn-in requirement. It also translates to a full-range minimax-optimal PAC sample complexity up to log factors.

Further, we extend our theory to unveil the influences of problem-dependent quantities like the optimal value/cost and certain variances.

The key technical innovation lies in a novel analysis paradigm to decouple complicated statistical dependency --- a long-standing challenge facing the analysis of online RL in sample-hungry scenarios.

3 - Speeding up Non-Parametric Regression via Distribution Compression

Albert Gong, Cornell University, Ithaca, NY, United States, Kyuseong Choi, Raaz Dwivedi

Modern compression methods have found diverse applications in speeding up integration, non-parametric hypothesis testing, and MCMC simulation. However, it remains unclear how to extend these recent advancements from the unsupervised learning domain to the supervised learning domain. We introduce a meta-algorithm based on Kernel Thinning that covers non-parametric regression problems and applies to general classes of kernel functions. In particular, we investigate two classical algorithms, namely the Nadaraya-Watson (NW) estimator and the Kernel Ridge Regression (KRR) estimator, for which we construct two regression meta-kernels. We show how generalization in each setting reduces to constructing the appropriate kernel and introduce the Kernel-Thinned NW and Kernel-Thinned KRR estimators, which enjoy improved computational efficiency over the vanilla estimators and improved statistical efficiency over i.i.d. subsampling of the training data. We derive excess risk rates for both estimators and validate our design choices with experiments on simulated and real-world data.

4 - Statistical Methods for Transfer Learning: Prediction Under Covariate Shift

Reese Pathak, UC Berkeley, Berkeley, CA, United States

Contemporary predictive settings pose challenges for traditional statistical methodology, partly due to distribution shifts between training and test data. These distributional mismatches are known to cause dramatic reductions in model performance. How can we optimally address these challenging predictive settings? In this talk, we study the fundamental limits of covariate shift: problems where the feature distribution changes between the train and test data. We focus on high-dimensional and nonparametric regression, revealing multiple surprises regarding the success and failures of classical methods. We introduce new measures of "shift" which are adapted to both the geometry induced by the underlying model class as well as the pair of train and test distributions. We complement upper bounds on model performance under covariate shift with lower bounds that showcase the optimality of our methods and demonstrate that our new notions can characterize the difficulty of the covariate shift problem at hand.

Based on joint work with Martin J. Wainwright (MIT), Cong Ma (Univ. of Chicago), and Lin Xiao (Meta).

SD43

Summit - 435

Strategic and Distributionally Robust Sequential Decision Making

Invited Session

Applied Probability Society

Chair: Adam Wierman, California Institute of Technology, Pasadena, CA, United States

Co-Chair: Kishan Panaganti, California Institute of Technology, MC 305-16 2100, Pasadena, 91106, United States

Co-Chair: Laixi Shi, California Institute of Technology (Caltech), Pasadena, CA, United States

1 - Model-Free Robust Φ-Divergence Reinforcement Learning Using Both Offline and Online Data

Kishan Panaganti, California Institute of Technology, Pasadena, CA, United States, Adam Wierman, Eric Mazumdar

The robust φ-regularized Markov Decision Process (RRMDP) framework focuses on designing control policies that are robust against parameter uncertainties due to mismatches between the simulator (nominal) model and real-world settings. This work makes *two* important contributions. First, we propose a *model-free* algorithm called *Robust φ-regularized fitted Q-iteration* (RPQ) for learning an ε-optimal robust policy that uses only the historical data collected by rolling out a behavior policy (with *robust exploratory* requirement) on the nominal model. To the best of our knowledge, we provide the *first* unified analysis for a class of φ-divergences achieving robust optimal policies in high-dimensional systems with general function approximation. Second, we introduce the *hybrid robust φ-regularized reinforcement learning* framework to learn an optimal robust policy using both historical data and online sampling. Towards this framework, we propose a model-free algorithm called *Hybrid robust Total-variation-regularized Q-iteration* (HyTQ: pronounced *height-Q*). To the best of our knowledge, we provide the *first* improved out-of-data-distribution assumption in large-scale problems with general function approximation under the hybrid robust φ-regularized reinforcement learning framework. Finally, we provide theoretical guarantees on the performance of the learned policies of our algorithms on systems with arbitrary large state space.

2 - Minimax Optimal and Computationally Efficient Algorithms for Distributionally Robust Offline Reinforcement Learning

Pan Xu, Duke University, Durham, NC, United States

Distributionally robust offline reinforcement learning (RL), which seeks robust policy training against environment perturbation by modeling dynamics uncertainty, calls for function approximations when facing large state-action spaces. However, the consideration of dynamics uncertainty introduces essential nonlinearity and computational burden, posing unique challenges for analyzing and practically employing function approximation. Focusing on a basic setting where the nominal model and perturbed models are linearly parameterized, we propose minimax optimal and computationally efficient algorithms realizing function approximation and initiate the study on instance-dependent suboptimality analysis in the context of robust offline RL. Our results uncover that function approximation in robust offline RL is essentially distinct from and probably harder than that in standard offline RL. Our algorithms and theoretical results crucially depend on a variety of new techniques, involving a novel function approximation mechanism incorporating variance information, a new procedure of suboptimality and estimation uncertainty decomposition, a quantification of the robust value function shrinkage, and a meticulously designed family of hard instances, which might be of independent interest.

3 - Wasserstein Distributionally Robust Policy Learning with Continuous Context

Wenhao Yang, Stanford University, Stanford, CA, United States, Miao Lu, Zhengyuan Zhou

In the offline contextual bandit problem, the optimal policy derived from an offline dataset often proves to be sensitive to changes in the environment. To handle this challenge, distributionally robust optimization (DRO) techniques are introduced to policy learning to enhance the robustness of the learned policy. However, most existing results only study the case when the context space is discrete. In this paper, we investigate the Wasserstein distributionally robust policy learning problem when the context space is continuous. We first propose a theoretical algorithm to solve this problem using the Nadaraya–Watson estimator in non-parametric statistics, and we establish both non-asymptotic and asymptotic theory to show the effectiveness of the method. To our best knowledge, this study gives the first theoretical result for Wasserstein distributionally robust policy learning problem with continuous context. Furthermore, we develop a practical algorithm to tackle the computational challenges to implement our theoretical algorithm. We evaluate its performance in the application of a contextual newsvendor problem.

4 - Sample-Efficient Robust Multi-Agent Reinforcement Learning in the Face of Environmental Uncertainty

Laixi Shi, California Institute of Technology (Caltech), Pasadena, CA, United States

To overcome the sim-to-real gap in reinforcement learning (RL), learned policies must maintain robustness against environmental uncertainties. While robust RL has been widely studied in single-agent regimes, in multi-agent environments, the problem remains understudied—despite the fact that the problems posed by environmental uncertainties are often exacerbated by strategic interactions. This work focuses on learning in distributionally robust Markov games (RMGs), a robust variant of standard Markov games, wherein each agent aims to learn a policy that maximizes its own worst-case performance when the deployed environment deviates within its own prescribed uncertainty set. This results in a set of robust equilibrium strategies for all agents that align with classic notions of game-theoretic equilibria. Assuming a non-adaptive sampling mechanism from a generative model, we propose a sample-efficient model-based algorithm (DR-NVI) with finite-sample complexity guarantees for learning robust variants of various notions of game-theoretic equilibria. We also establish an information-theoretic lower bound for solving RMGs, which confirms the near-optimal sample complexity of DR-NVI with respect to problem- dependent factors such as the size of the state space, the target accuracy, and the horizon length.

SD44

Summit - 436

Recent Advances in Stochastic Simulation

Invited Session

Simulation Society

Chair: Dohyun Ahn, CUHK, Shatin, Hong Kong

1 - Gradient Estimation of Stationary Mean and Average Reward Reinforcement Learning

Jeffrey Wang, Northwestern University, Chicago, IL, United States, Chang-Han Rhee

First part of our work focuses on developing unbiased and efficient estimators for the gradient of steady-state mean of a parametrized Markov chain. Then we extend our results to the sequential optimization setting, specifically, reinforcement learning (RL) with long-run average reward objective. Previously developed trust region based RL methods have shown great performance for the discounted reward setting RL problems but naive implementation of those methods in the average reward setting often fails due to instability caused by the lacking of a discounting factor. We extend our results in the stationary mean gradient estimation section to this setting and develop a multi-level average reward trust region based RL algorithm that is more efficient and more importantly, more stable.

2 - Locally Robust EstimatiOn for Mitigating Input Uncertainty in Minimization of Tail Risks

Anand Deo, Indian Institute of Management Bangalore, Bangalore, India, Karthyek Murthy, Arjun Ramachandra

The ability to learn and control tail risks, besides being an integral part of quantitative risk management, is central to running operations requiring high service levels and cyber-physical systems with high-reliability specifications. Due to the paucity of relevant samples in the tail regions, formulations involving tail risks are almost always approached with the "estimate, then optimize" workflow involving a model estimation from data in the first step before plugging in the trained model to solve various downstream decision-making tasks via simulation. As biases due to model selection, misspecification, and overfitting to in-sample data are difficult to avoid in the first-step estimation, we construct novel locally robust estimators in which the input uncertainty due to the first-step estimation has no effect, locally, on the decisions obtained via downstream simulation. We show that this *local insensitivity* translates to improved out-of-sample performance freed from the first-order impact of input model uncertainty introduced in the first-step estimation.

A key ingredient in achieving this local robustness is a novel debiasing procedure that adds a non-parametric bias correction term to the objective. The debiased formulation retains convexity, and the imputation of the correction term relies only on a non-restrictive large deviations behavior conducive for transferring knowledge from representative data-rich regions to the data-scarce tail regions. The bias correction gets determined by the extent of model error in the estimation step and the specifics of the decision-making task in the optimization step, thereby serving as a scalable "smart-correction" step bridging the disparate goals in estimation and optimization.

3 - Generalized Importance Sampling for Nested Simulation

Mingbin Feng, University of Waterloo, Waterloo, ON, Canada, Amber (Qingyuan) Chen

Importance sampling (IS) is a classical variance reduction technique. Under mild conditions, an IS estimator is unbiased, so one often seeks variance-minimizing optimal sampling distribution. IS has remarkable success in many applications such as engineering, operations research, and finance. In some applications such as enterprise risk management and input uncertainty quantification, complex simulation designs such as nested simulation arises naturally: The outer-level simulation generates a set of risk factors, i.e., the scenarios, which are used as inputs for inner-level simulations. Nested simulation leads to wasteful use of computations as inner simulation outputs in each scenario are isolated from other scenarios. In this study, we propose, analyze, and test a generalized importance sampling technique for nested simulation. Our generalized IS approach reuses one set of inner simulation outputs across different outer scenarios. Numerical experiments show that our proposal is orders of magnitudes more efficient than the standard procedure.

4 - Data-Driven Sequential Sampling for Tail Risk Mitigation

Dohyun Ahn, The Chinese University of Hong Kong, Shatin, Hong Kong, Taeho Kim

Given a collection of stochastic alternatives, we study the problem of sequentially allocating a fixed sampling budget to identify the optimal alternative with a high probability, where the optimal alternative is defined as the one with the smallest value of extreme tail risk. We particularly consider a situation where these alternatives generate heavy-tailed losses whose probability distributions are unknown and may not admit any specific parametric representation. In this setup, we propose a data-driven sequential allocation rule that maximizes the rate at which the likelihood of falsely selecting suboptimal alternatives decays to zero. We rigorously demonstrate the superiority of the proposed method over existing approaches, which is further validated via numerical studies.

5 - Queuing Systems Control via Reinforcement Learning

Gabriel Nicolosi, Missouri University of Science and Technology, Rolla, MO, United States

With the proliferation of machine learning models into every area of operations research and the management sciences, the problem of stochastic queuing systems control is being reconsidered from the perspective of reinforcement learning. Industrial, economic and human-centered systems of current practical interest have gained such an intricate complexity that the established tools provided by the classical apparatus of dynamic programming are no longer suitable for their optimization and control, both from modeling and computational perspectives. Moreover, many of these systems are becoming cyber-physical, designed alongside their so-called digital twins. This new digital paradigm allows for a feedback loop between the physical system and its digital counterpart, wherein simulation-based techniques, such as reinforcement learning, gain an important role. This approach might prove beneficial in the optimization and control of world-critical systems where congestion appears, such as semiconductor manufacturing, traffic control in smart cities and healthcare supply chains. In this presentation, the reinforcement learning-based control of a M/G/k queuing system under server breakdowns is considered. A concise overview of the field will be presented. Modeling, algorithmic and implementation aspects will be discussed, followed by current computational results.

SD45

Summit - 437

Broadening the Reach of OR for Healthcare: Publishing in Non-OR/IE Journals

Panel Session

Health Applications Society

Co-Chair: Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Panelist

Brian Denton, University of Michigan, Ann Arbor, MI, United States

2 - Panelist

Oguzhan Alagoz, University of Wisconsin-Madison, Madison, WI, United States

3 - Panelist

Margaret Brandeau, Stanford University, Stanford, CA, United States

4 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

1

SD46

Summit - 438

Featured Journal - Health Care Management Science

Panel Session

Health Applications Society

Co-Chair: Greg Zaric, Ivey Business School, London, ON, Canada

1 - Moderator Panelist

Greg Zaric, Ivey Business School, London, ON, Canada

2 - Panelist

Yuan Shi, MIT, Boston, MA, United States, Saied Mahdian, Jose Blanchet, Peter Glynn, David Scheinker

Surgical scheduling via optimization and machine learning with long-tailed data. https://link.springer.com/article/10.1007/s10729-023-09649-0 Using data from cardiovascular surgery patients with long and highly variable post-surgical lengths of stay (LOS), we develop a modeling framework to reduce recovery unit congestion. We estimate the LOS and its probability distribution using machine learning models, schedule procedures on a rolling basis using a variety of optimization models, and estimate performance with simulation. The machine learning models achieved only modest LOS prediction accuracy, despite access to a very rich set of patient characteristics. Compared to the current paper-based system used in the hospital, most optimization models failed to reduce congestion without increasing wait times for surgery. A conservative stochastic optimization with sufficient sampling to capture the long tail of the LOS distribution outperformed the current manual process and other stochastic and robust optimization approaches. These results highlight the perils of using oversimplified distributional models of LOS for scheduling procedures and the importance of using optimization methods well-suited to dealing with long-tailed behavior.

3 - Panelist

Soroush Saghafian, Harvard University, Medford, CT, United States, Mariam Atkinson

Who should see the patient? on deviations from preferred patient-provider assignments in hospitals. https://link.springer.com/article/10.1007/s10729-022-09628-x In various organizations including hospitals, individuals are not forced to follow specific assignments, and thus, deviations from preferred task assignments are common. This is due to the conventional wisdom that professionals should be given the flexibility to deviate from preferred assignments as needed. It is unclear, however, whether and when this conventional wisdom is true. We use evidence on the assignments of generalist and specialists to patients in our partner hospital (a children's hospital), and generate insights into whether and when hospital administrators should disallow such flexibility. We do so by identifying 73 top medical diagnoses and using detailed patient-level electronic medical record (EMR) data of more than 4,700 hospitalizations. In parallel, we conduct a survey of medical experts and utilized it to identify the preferred provider type that should have been assigned to each patient. Using these two sources of data, we examine the consequence of deviations from preferred provider assignments on three sets of performance measures: operational efficiency (measured by length of stay), quality of care (measured by 30-day readmissions and adverse events), and cost (measured by total charges). We find that deviating from preferred assignments is beneficial for task types (patients' diagnosis in our setting) that are either (a) well-defined (improving operational efficiency and costs), or (b) require high contact (improving costs and adverse events, though at the expense of lower operational efficiency). For other task types (e.g., highly complex or resource-intensive tasks), we observe that deviations are either detrimental or yield no tangible benefits, and thus, hospitals should try to eliminate them (e.g., by developing and enforcing assignment guidelines). To understand the causal mechanism behind our results, we make use of mediation analysis and find that utilizing advanced imaging (e.g., MRIs, CT scans, or nuclear radiology) plays an important role in how deviations impact performance outcomes. Our findings also provide evidence for a "no free lunch" theorem: while for some task types, deviations are beneficial for certain performance outcomes, they can simultaneously degrade performance in terms of other dimensions. To provide clear recommendations for hospital administrators, we also consider counterfactual scenarios corresponding to imposing the preferred assignments fully or partially, and perform cost-effectiveness analyses. Our results indicate that enforcing the preferred assignments either for all tasks or only for resource-intensive tasks is cost-effective, with the latter being the superior policy. Finally, by comparing deviations during weekdays and weekends, early shifts and late shifts, and high congestion and low congestion periods, our results shed light on some environmental conditions under which deviations occur more in practice.

4 - Panelist

Christina Bartenschlager, Ohm University of Applied Sciences Nuremberg, Nürnberg, Germany, Milena Grieger

Covid-19 triage in the emergency department 2.0: how analytics and AI transform a human-made algorithm for the prediction of clinical pathways. The Covid-19 pandemic has pushed many hospitals to their capacity limits. Therefore, a triage of patients has been discussed controversially primarily through an ethical perspective. The term triage contains many aspects such as urgency of treatment, severity of the disease and pre-existing conditions, access to critical care, or the classification of patients regarding subsequent clinical pathways starting from the emergency department. The determination of the pathways is important not only for patient care, but also for capacity planning in hospitals. We examine the performance of a human-made triage algorithm for clinical pathways which is considered a guideline for emergency departments in Germany based on a large multicenter dataset with over 4,000 European Covid-19 patients from the LEOSS registry. We find an accuracy of 28 percent and approximately 15 percent sensitivity for the ward class. The results serve as a benchmark for our extensions including an additional category of palliative care as a new label, analytics, AI, XAI, and interactive techniques. We find significant potential of analytics and AI in Covid-19 triage regarding accuracy, sensitivity, and other performance metrics whilst our interactive human-AI algorithm shows superior performance with approximately 73 percent accuracy and up to 76 percent sensitivity. The results are independent of the data preparation process regarding the imputation of missing values or grouping of comorbidities. In addition, we find that the consideration of an additional label palliative care does not improve the results.

5 - Panelist

Daniel Otero Leon, Harvard Medical School, Boston, MA, United States

Monitoring policy in the context of preventive treatment of cardiovascular disease. https://link.springer.com/article/10.1007/s10729-022-09621-4 Preventing chronic diseases is an essential aspect of medical care. To prevent chronic diseases, physicians focus on monitoring their risk factors and prescribing the necessary medication. The optimal monitoring policy depends on the patient's risk factors and demographics. Monitoring too frequently may be unnecessary and costly; on the other hand, monitoring the patient infrequently means the patient may forgo needed treatment and experience adverse events related to the disease. We propose a finite horizon and finite-state Markov decision process to define monitoring policies. To build our Markov decision process, we estimate stochastic models based on longitudinal observational data from electronic health records for a large cohort of patients seen in the national U.S. Veterans Affairs health system. We use our model to study policies for whether or when to assess the need for cholesterol-lowering medications. We further use our model to investigate the role of gender and race on optimal monitoring policies.

6 - Panelist

Dmitry Krass, Rotman School of Management, University of Toronto, Toronto, ON, Canada

Machine learning for optimal test admission in the presence of resource constraints. https://link.springer.com/article/10.1007/s10729-022-09624-1 Developing rapid tools for early detection of viral infection is crucial for pandemic containment. This is particularly crucial when testing resources are constrained and/or there are significant delays until the test results are available – as was quite common in the early days of Covid-19 pandemic. We show how predictive analytics methods using machine learning algorithms can be combined with optimal pre-test screening mechanisms, greatly increasing test efficiency (i.e., rate of true positives identified per test), as well as to allow doctors to initiate treatment before the test results are available. Our optimal test admission policies account for imperfect accuracy of both the medical test and the model prediction mechanism. We derive the accuracy required for the optimized admission policies to be effective. We also show how our policies can be extended to re-testing high-risk patients, as well as combined with pool testing approaches. We illustrate our techniques by applying them to a large data reported by the Israeli Ministry of Health for RT-PCR tests from March to September 2020. Our results demonstrate that in the context of the Covid-19 pandemic a pre-test probability screening tool with conventional RT-PCR testing could have potentially increased efficiency by several times, compared to random admission control.

SD48

Summit - 440

AI and Information Technology in Operations Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Jingjing Weng, Temple University, Philadelphia, PA, United States

1 - Bi-Level Optimization in a Collaborative Supply Chain: Production and Procurement Planning Based on Suppliers' Production Decisions

Ya Zhou, Hefei University of Technology, Hefei, China, People's Republic of, Jun Pei

To leverage the tremendous value of supply chain collaboration, manufacturers embrace suppliers as partners with whom to undertake collaborative supply chain planning that considers suppliers' profits and decisions. This study formulates the collaborative supply chain planning problem as a bi-level optimization problem. At the upper level, the manufacturer makes the parts procurement and products production planning with the aim of minimizing the total costs by anticipating suppliers' responses to parts procurement orders. At the lower level, the suppliers make their respective part production planning, including decisions on order acceptance, production line opening, and assignment of the accepted parts. Considering the mutual interactions between the decisions at the two levels, we develop a sequential reduction algorithm to implement the nested decomposition of the studied bi-level problem. Under the proposed reduction algorithm framework, the lower-level problem is first solved by an improved branch-and-bound-based (B&B-based) reduction algorithm with the double pruning operations and enhanced mathematical model. Then, the bi-level problem is reduced through nesting the optimal solutions for lower-level problems into the overall solution, i.e., hierarchical fusion reduction. Finally, a novel variable neighborhood search-based (VNS-based) reduction algorithm is utilized to get near-optimal solutions for the reduced bi-level problem. Computational experiments are conducted to verify the performance of the sequential reduction algorithm framework. Moreover, based on the experiment results, this study provides insightful observations on sourcing diversification and decision preferences of large manufacturing enterprises.

2 - Task Flexibility in Gig Platforms

Zhi Cao, Sichuan University, Chengdu, China, People's Republic of, Meng Li, Shujing Sun

Flexibility is a hallmark of gig-economy platforms, yet how it influences gig workers' behaviors remains underexplored, especially in the healthcare industry. Leveraging a quasi-natural experiment on a gig healthcare platform, this paper investigates the impact of task flexibility on individuals' behaviors and platform sustainability. Using the different-in-differences design, we find that the decline in task flexibility significantly decreases individuals' willingness to engage in tasks and efforts in service delivery. We further uncover the underlying mechanism through an experiment. The results show that the negative effects are attributable to changes in individuals ' psychological perceptions, particularly feelings of unfairness, disrespect, and low economic expectations stemming from the reduced task flexibility. Our findings provide novel insights into the flexibility design for gig platforms.

3 - Managing Cybersecurity: Data Access & Protection

Oleh Stupak, University of Cambridge, Cambridge, United Kingdom, George Charlson, Ruslan Momot, Marat Salikhov

A standard data access guideline is to limit employee access to only the essential data for their roles, raising the question of how to define these roles and the employees' corresponding data needs. We aim to address this question by considering a game-theoretic model of joint cybersecurity and operational decision-making. The firm chooses the level of data access for each of its employees and the overall level of cyber protection. Providing more data access to employees makes the firm more economically efficient but also more vulnerable to attacks by an adversary who steals the employees' data, inflicting damage that increases with the amount of data stolen. Adversaries may vary in their attack strength and sophistication rate (the ability to pinpoint the most attractive targets). We find that the firm may counter-intuitively decrease its overall protection level when adversarial attacks become stronger and increase its overall access level when the adversaries become more sophisticated.

SD49

Summit - 441

Platform and AI

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Rachel Chen, University of California-Davis, 3208 Gallagher Hall, Davis, CA, 95616, United States

Co-Chair: Xuying Zhao, Texas A&M University, College Station, TX, United States

1 - Revenue Sharing with Swap: Application in Digital Product Development

Xuying Zhao, Texas A&M University, College Station, TX, United States, Leon xu, Jingmiao Song

In the context of digital product development, where developers contribute creative content and publishers provide funding and marketing support, traditional revenue-sharing contracts often raise fairness concerns. This is because the dominating party usually claims a larger share of the revenue, thus discouraging the other party from exerting its best effort. To address this concern, the video game industry has introduced an innovative revenue-sharing contract with swap: the favored party will be swapped (sometimes with a different revenue-split ratio) once the publisher recovers its investment. However, it remains unclear whether such a swap contract will be mutually beneficial. Therefore, we investigate a Stackelberg game between the developer and publisher, examining three types of revenue-sharing contracts: no-swap, swap, and reciprocal swap (where the favored parties are swapped but the revenue-split ratio remains unchanged). Compared to the no-swap contract, we find that the swap contract incentivizes the publisher's investment and the developer's effort more effectively. However, the publisher may use the swap contract's inherent flexibility to its advantage, benefiting unilaterally from the contract. Only when development requires significant funding does the swap contract benefit both parties mutually. Interestingly, the reciprocal swap contract not only upholds intrinsic fairness merit but also achieves most of the optimal supply chain profit, thereby better coordinating the supply chain than both the swap and no-swap contracts. Our study provides valuable insights for developers negotiating contracts with publishers and sheds light on revenue-sharing contracts when contributions from different parties are complementary.

2 - Managing Trade-Offs Between Revenue, Access and Market Balance on Online Dating Platforms

Qiqi Hao, University of Pittsburgh, Pittsburgh, PA, United States, Titing Cui, Michael Hamilton

Online dating platforms, like many other online platforms, generate revenue by charging users for enhanced services. However, unlike other online services, for dating platforms to function properly they must maintain a large pool of users and, for heterosexual dating, maintain approximate parity between the number of male and female users. If the platform fails to maintain a large candidate pool or balance the two sides of their market, the number of matches made on the platform can plummet and the platform's value proposition can collapse. The primary lever under a platform's control to achieve these goals is segmented pricing, which, in practice, leverages a combination of geographic and demographic features to set semi-personalized prices for subsets of users.

In this study, we explore the trade-offs between revenue, access, and market balance for dating apps through the lens of segmented pricing. We present a novel model that formalizes the dating platform's problem as a non-convex mathematical program whose objective is a linear combination of the three market outcomes. Under natural assumptions we show how to solve this optimization problem and examine the resulting market outcomes. As an application of our framework, we use our model and solution to investigate regulatory calls for enforcing price and access fairness on dating platforms. We encode these proposed regulations as constraints on the platforms pricing problem and explore the potential pros and cons associated with the new outcomes they induce.

3 - The Economics of AI Foundation Models: Openness, Competition, and Governance

Fasheng Xu, University of Connecticut, Stamford, CT, United States, Wei Chen, Xiaoyu Wang, Karen Xie

AI is undergoing a paradigm shift with the rise of foundation models (e.g., GPT-4, Claude 3, Gemini, Llama 2, Stable Diffusion) trained on broad data using self-supervision at immense scale, which can then be adapted to myriad downstream tasks. This paper offers an economic theory of foundation model ecosystems that consist of upstream model developers and downstream model deployers. We model how fine-tuning games arise as downstream deployers competing in foundation model adoption and fine-tuning effort, and how foundation model openness affects the fine-tuning games. Our theory provides insights into the economic implications and trade-offs for multiple stakeholders throughout the AI value chain (including developer, deployers, and consumers) and sheds light on how to harness the full potential of the foundation model value chain interactions and avoid pitfalls. An important implication of our findings is that the policymakers should explicitly consider potential unintended economic consequences of AI regulation on the ecosystem around open foundation models, in particular, the so-called ``openness trap'' (i.e., a range of medium openness levels that should be avoided). Furthermore, we explore the welfare implications of prevalent market strategies employed by upstream developers, such as vertical integration and offering free trials. Our findings reveal that vertical integration proves effective when model openness is relatively limited. The developer's strategy of providing free trials can negatively affect the leading deployer within a moderate range of model openness, while benefiting all other stakeholders.

4 - Genai Assistance in a Professional Service Market: the Perish of Second Opinion

Rachel Chen, University of California-Davis, Davis, CA, United States, Jane Gu, Yi Liu

We investigate the impact of generative AI (GenAI) adoption in a market of professional services such as medical treatment and financial consulting, where the product is a professional opinion. Considering a vertically differentiated market of professional services where a high-expertise provider and two low-expertise providers serve consumers with heterogeneous quality preferences, our study reveals important strategic implications of AI adoption. First, we show that after service providers adopt AI assistance, the high-expertise provider may enhance profit, despite quality improvement occurring to its low-expertise rivals but not to itself. In contrast, the low-expertise providers always suffer a severe drop in margins, although their total payoffs may increase under a sufficiently large benefit in corporate image associated with AI adoption. Two interesting regions arise: a ``lose-lose'' region, where AI adoption harms all service providers, and a ``win-lose'' region, where the high-expertise provider is better off but neither of the low-expertise providers. Second, we show that after providers' AI adoption, low-type consumers always enjoy a surplus gain but high-type consumers may suffer a surplus loss. Total consumer surplus declines if low-expertise providers are highly differentiated in opinions before AI adoption. Lastly, when the two low-expertise providers make endogenous decisions regarding whether to adopt AI assistance, equilibrium may arise where neither of them adopt AI assistance or only one adopts.

SD50

Summit - 442

Public Health Decision-Making for Infectious Diseases

Invited Session

MSOM: Healthcare

Chair: Sarang Deo, Indian School of Business, Hyderabad, India

Co-Chair: Pankaj Jindal, Indian School of Business (ISB), N/A

1 - Improving Decision-Making in Public Health Surveillance of Epidemics with Evolving Pathogens

Pankaj Jindal, Indian School of Business, Hyderabad, India, Sarang Deo, Sripad Devalkar

A key challenge to effective management of epidemics is rapid detection of novel variants which allows for implementing appropriate public health interventions. However, as demonstrated during the COVID-19 pandemic, capacity of the genomic sequencing-based surveillance systems, used for detecting novel variants, is often limited (particularly in LMICs) leading to delayed detection of the novel variants. It is thus crucial to study the impact of the design of genomic surveillance systems on the delay in detection of novel variants and effectiveness of public health interventions.

In this paper, we develop an analytical model, complemented by extensive numerical analysis, to obtain insights regarding the impact of operational characteristics such as sampling proportion and sequencing capacity on the delay in detection. We find that sequencing a higher proportion of diagnosed samples is not always optimal. This is because a higher sampling proportion can delay detection due to longer wait times at the sequencing lab due to congestion, even though it increases the probability of selecting a sample infected with the novel variant for sequencing. We also characterize the behavior of the optimal sampling proportion with other system parameters. We find that the optimal sampling proportion increases with increase in cost of non-detection and sequencing capacity and decreases with increase in holding cost and diagnosis rate.

Our results have practical implications in public health decision making for effective management of evolving epidemics and can lead to more efficient provision and utilization of genomic sequencing capacity and public health interventions.

2 - Using Simulation Modeling to Inform Strategies for the Eradication of Guinea Worm Disease

Hannah Smalley, ISyE Georgia Tech, Atlanta, GA, United States, Pinar Keskinocak, Maryann Delea, Obiora Eneanya, Adam Weiss

In 1986, there were an estimated 3.5 million human cases of Guinea worm disease (GWD) around the world. Since then, GWD has been eliminated in 17 countries, with only 14 human cases globally in 2023. However, in the same year, 407 dog infections were reported across 277 villages in Chad. Because Guinea worm (GW) can be transmitted through shared water sources, the human population remains at risk. Current interventions include tethering dogs with emerging worms and water treatment to reduce the number of infective copepods in water sources. Proactive tethering of dogs is employed to minimize exposure of dogs to potentially infected water sources, but household adherence to tethering guidelines is not always uniform or ideal, with some dogs tethered only intermittently or released at night. Tests to diagnose GW infection in dogs before signs of infection are under development. We built an agent-based simulation which models the transmission of GW among dogs in Chad to analyze intervention strategies. We find that tethering and water treatment levels need to increase to achieve GW elimination in Chad. Releasing dogs from proactive tethering for part of the day results in up to 60 times more infections after five years compared to full day tethering. A diagnostic test for prepatent infection could be helpful and support elimination efforts by decreasing time to elimination to as little as 2 years, but achieving this result relies on the implementation strategy and education as 2-year elimination requires high long-term tethering adherence following a positive test result.

3 - Pooled Testing in the Presence of Congestion

Oualid Jouini, CentraleSupelec, Gif-sur-Yvette, France, Saif Benjaafar, Benjamin Legros, Bingnan Lu

Pooled testing for infectious diseases has a rich history. However, the focus of much of the academic literature has been on maximizing testing throughput. While this approach maximizes the efficiency of the testing facility, it may not minimize the delay in obtaining test results. In this study, we provide a procedure, based on a queueing model, for choosing sample batch sizes that minimize expected delay. We show that not doing so, and relying on throughput maximization, may lead to excessive delays.

4 - Push OR Pull? -- when and How to Fund Drug Development

Peng Sun, Duke University, Durham, NC, United States, Chenxi Xu, David Ridley

Rare disease drugs often do not generate sufficient profit from the market to cover the development costs. In this case, pharmaceutical firms do not have the incentive to develop these drugs, unless receiving investments from external funding agencies. However, pharmaceutical firms often possesses much more information about the potential drug and the development processes, causing adverse selection and moral hazard issues. In order to use their resources efficiently, a funding agency needs to decide when and how to fund a pharmaceutical firm and manage the aforementioned incentive issues. For example, investing early (push) may save money, but does not offer accountability compared with paying after resolving development uncertainties (pull). We formulate the funding agency's problem as a mechanism design model, and theoretically study optimal mechanisms under different settings. An interesting finding is that it is sometimes optimal for the funding agency to reward failure.

SD51

Summit - 443

Financing Operations

Invited Session

MSOM: iForm

Chair: Jiri Chod, Boston College, Chestnut Hill, MA, United States

1 - On the Benefits of Supplier Vs Customer Diversification: a Geographical Analysis

Deepak Agrawal, Emory University, Atlanta, GA, United States, Nikolay Osadchiy

Diversification is traditionally thought to be a risk mitigation tool. However, in our paper, we show that diversification also have first order effect on inventory productivity. We find a nonlinear relationship between geographical diversification and inventory turnover. Further, we examine the varying effects of supplier vs customer diversification by firm's upstreamness, competitive environment, and innovation.

2 - On the Interplay of Production Flexibility and Financing Strategy

Guoming Lai, The University of Texas at Austin, Austin, TX, United States, Peter Ritchken, Qi Wu

In this paper, we explore how the improved production flexibility impacts joint operating policies and financing decisions. We develop a dynamic model, in which the equity holders of the firm make the decisions of pausing and restarting of production, and the degree of flexibility is reflected by the switching cost. We find that the optimal operating policy is jointly determined with the financing choice. Debt levels, not only impact tax shields and bankruptcy costs but also the utilization of production flexibility. Debt induces risk shifting which undermines equity holders' incentive to use flexibility. We further uncover a non-monotone relationship between production flexibility and financial leverage. When the switching cost is low, production flexibility complements the benefits of debt and thus financial leverage increases in production flexibility. As the switching cost increases into an intermediate region, the firm aggressively reduces its debt to ensure the continued usage of flexibility. However, when the switching cost exceeds a threshold, the cost of reducing leverage to maintain production flexibility becomes too high, and the firm forgoes flexibility and establishes high financial leverage.

3 - Risk Mitigation Role of Trade Credit for a Dual-Channel Supplier: Theory and Evidence

Jie Ning, Case Western Reserve University, Cleveland, OH, United States, Sammi Tang, Jing Wu, Jiaqi Zhu

Trade credit is commonly perceived to exacerbate supplier risk, as delayed payments expose the supplier to buyer's uncertainties. In contrast, cash on delivery (COD) is viewed safe, as it insulates the supplier from downstream risk. Using a stylized game theoretic model, we show that this conventional wisdom may be overturned for a dual-channel supplier who operates a direct channel that sells to customers, in addition to an indirect channel that sells to a buyer who then sells to customers. When the supplier and buyer's direct sales are *negatively correlated* due to their market competition or channel risks, trade credit yields two negatively correlated payment streams for the supplier, which enables risk pooling and can yield *lower* supplier risk than COD. Interestingly, the buyer also benefits and earns higher expected payoff than under COD, as the supplier lowers the wholesale price to achieve a more "balanced" portfolio to enhance pooling. We empirically validate the risk mitigation effect of trade credit using data on manufacturing firms that sell to their competitors. We show that a supplier's short-term leverage and stock price volatility *decrease* with its accounts receivable after selling to competitors, while being invariable in accounts receivable before such transactions. These effects are concentrated on suppliers that have negatively correlated sales or high competition intensity with their buyers.

4 - Interaction of Operations Decision and Financial Derivatives: An Empirical Perspective

Gangshu Cai, Santa Clara University, Santa Clara, CA, United States, Wubo Zhang, QI ZHANG

This study aims to contribute to the underexplored area of how derivatives markets can bolster economic entities and supply chain growth. Using a sample of 289 chemical sector companies, this research employs empirical analysis with fixed effects and unsupervised machine learning to investigate firm motivations for adopting financial derivatives hedging from 2016 to 2022. Our findings reveal that operational metrics, notably inventory and turnover rates, play a significant role in influencing a company's hedging activities.

5 - Inventory, Renegotiation, and Trade Credit

Jiri Chod, Boston College, Chestnut Hill, MA, United States, Evgeny Lyandres

We study contract renegotiation between a seller and a financially constrained buyer. We compare efficiency of renegotiation and the initial trade agreement under bank financing and under trade credit financing. Our model uncovers a novel benefit of trade credit, which is the elimination of agency frictions during the bargaining process. The model also shows how this benefit depends on bargaining power distribution, level of uncertainty, and cost and revenue parameters of the good that is being traded.

SD52

Summit - 444

Stochastic Modeling in Energy Transition

Invited Session

MSOM: Service Operations

Chair: Yue Hu, Stanford University, Palo Alto, CA, United States

Co-Chair: Omer Karaduman, Stanford University, Stanford, United States

1 - Blood-Donation-Inspired Aggregator Business Model for Vehicle-to-grid Services

Yangfang (Helen) Zhou, Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore, Meichun Lin, Owen Wu, John Pang

We propose a novel business model to promote individual electric vehicle (EV) owners to adopt vehicle-to-grid (V2G), where electricity flows from EVs to the electric grid. The most critical concern hindering adoption is the potential degradation of the vehicle battery due to increased discharging, leading manufacturers to invalidate the battery warranty. To address this concern, we propose in our business model that each EV owner performs V2G services *a few times a year* during peak electricity demand. This model draws inspiration from blood donation practice: Paralleling blood donors contributing twice annually without compromising health, performing V2G services a limited number of times (e.g., five) ensures negligible battery degradation, alleviating concerns over the battery life expectancy for both EV owners and manufacturers. Under this business model, an aggregator enrols individual EV owners and coordinates them to transact collectively with the grid operator. Diverging from blood donation, the aggregator compensates EV owners for their V2G service.

We examine how to design such compensation schemes for both a non-profit aggregator and for-profit aggregator. We model the sequential game between the aggregator and the pool of EV drivers using a two-stage model: In the first stage, the aggregator offers a menu of compensation for each number of V2G service calls; in the second stage, given the chosen number of service calls of each EV driver, the aggregator solves a stochastic dynamic program to determine the number of drivers to call in the event of peak electricity demand.

2 - From Curtailed Renewable Energy to Green Hydrogen: Infrastructure Planning for Hydrogen Fuel-Cell Vehicles

Long He, George Washington University, Washington, DC, United States, Nan Ke, Ruijiu Mao, Wei Qi, Hongcai Zhang

Hydrogen fuel-cell vehicles (HFVs) offer a promising solution for green transportation, particularly in regions with renewable energy curtailment. This paper examines the challenges of promoting HFVs, including uncertain adoption rates, lack of infrastructure, and spatial mismatches between hydrogen demand and renewable sources. A planning model is developed to determine the location and capacities of hydrogen refueling stations (HRSs) and plants, as well as grid upgrades. Despite complexities like driver behavior, the model is tractable and applied to Sichuan, China, known for abundant hydro resources and curtailed hydropower. Key findings include: (i) Optimal HRS deployment varies with HFV adoption targets, with larger cities emerging as adoption centers at higher targets; (ii) HFV adoption can reduce hydropower curtailment, contingent on factors like adoption goals and grid upgrade costs; (iii) Hydrogen's transportability allows strategic HRS placement apart from plant sites, enhancing cost savings and curtailment reduction compared to electric vehicles.

3 - Conditional Quadratic Hedging of Term Structure Risk in Merchant Energy Trading Operations

Nicola Secomandi, Rice University, Houston, TX, United States, Bo Yang

We apply conditional quadratic hedging to the management of term structure risk in merchant energy trading operations. We develop a model with pooled cash flows across dates and establish the structure of its optimal policy, which we use to propose a computational efficient heuristic with theoretical appeal. This method performs near optimally in a realistic numerical study in the context of long duration merchant energy storage. Our approach has potential relevance to manage market price risk embedded in energy transition projects.

4 - Optimal Mechanisms for Demand Response in Energy Systems: A Mean-Field Approach

Mohammad Mehrabi, Stanford University, Stanford, CA, United States

The increased use of renewable energy resources has led to interest in demand response programs that can shift consumer demand to better match renewable production. We find that, in large systems and under mean-field asymptotics, pricing-based mechanisms can match the demand-response benefits achievable through direct optimization methods that use detailed knowledge of consumer preferences. In addition, we show that optimal prices can be efficiently learned using a simple first-order algorithm that only requires access to market-level metrics.

5 - Optimal Batch Scheduling for Power Grid Interconnection Queues

Omer Karaduman, Stanford University, Stanford, CA, United States, Yue Hu, Lin Zang

We model the interconnection process of connecting electricity generators to the U.S. power grid as a multi-stage queueing system. Developers submit their project proposals to the queue and undergo a multi-stage evaluation process. For this queueing system, we investigate various control levers related to capacity allocation and evaluation strategies from the perspective of the system manager. The goal is to optimize queueing-related performance metrics, such as reducing waiting times and improving system throughput.

SD53

Summit - 445

Data Analytics in Operations Management

Invited Session

MSOM: Supply Chain

Chair: Qi Feng, Purdue University, West Lafayette, IN, United States

Co-Chair: Jian Wu, Purdue University, West Lafayette, IN, United States

1 - Analytics for Supply Chain Resilience Strategies Using 3D-Printing Backup

Ziyu He, University of Southern California, Los Angeles, CA, United States, Vishal Gupta, Chamsi Hssaine, Nick Vyas

Conventionally, 3D printing (3DP) has been widely perceived as only adequate for limited roles in large supply chain systems, such as prototyping and producing spare parts, due to its lack of economies of scale. In this paper, we aim to expand the scope of 3DP applications by exploring its potential as a supply chain resilience strategy. Specifically, we examine a multi-product inventory system where 3DP backup is complemented by expedition services from traditional manufacturers (TM) to mitigate the risk of supplier disruptions. Here, 3DP is modeled as a flexible resource whose adoption cost does not scale with the number of protected products, albeit with a limited total recoverable capacity compared to TM expedition. To quantify the value of 3DP backup in enhancing the resilience of this system, we provide computational tools and insights on how to optimally build a backup strategy based on 3DP. This includes devising algorithms to compute optimal decisions such as which products should be protected by either backup sources and how to allocate the appropriate 3DP capacity to invest in. These numerical tools are further complemented by easy-to-implement heuristics that provide more interpretable insights into the drivers of backup policies and 3DP investments, such as marginal disruption probability and disruption correlations among suppliers.

2 - Inventory Commingling on An E-Commerce Platform

Hanqi Wen, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Izak Duenyas, Joline Uichanco

Amazon is an e-commerce marketplace that also provides fulfillment services for its third-party sellers. A recent strategy of Amazon is to commingle inventory, referring to pooling a third-party seller's inventory with the inventory of other sellers at its fulfillment centers. Inventory commingling benefits the platform by reducing fulfillment costs, as the commingled inventory are interchangeable regardless of which seller makes the sale. For sellers, commingling inventory can reduce the cost it would otherwise incur of labeling its inventory with a seller-specific sticker. However, the seller faces the risk that the product sold under her name may be a counterfeit/defective product provided by another seller. In this talk, we study the marketplace dynamics of inventory commingling through a model where sellers with heterogeneous product qualities are choosing whether or not to participate in inventory commingling. We find that this decision to "free ride" versus protect their competitive advantage depends on the product quality, pre-handling cost, product revenue, and fulfillment fees. We design mechanisms for the platform to maximize profits. When the platform generates revenue through both commission fees and fulfillment fees, complicated trade-offs happen. By appropriately deciding the commission rate and fulfillment fee, the platform can leverage inventory commingling to its fullest advantage.

3 - Deep Learning for Visual Advertising on Digital Platforms: Asymptotically Optimal Image Selection

Bora Keskin, Duke University, Durham, NC, United States, Yuexing Li, Shaoxuan Liu, Jing-Sheng Song

We consider a digital platform that aims to crop and display a large number of images to maximize customer conversions, i.e., customer purchasing decisions. For each image, the platform chooses a cropping window and observes the resulting conversions. The platform does not know how cropped images influence conversions. We design a novel two-stage deep-learning policy that dynamically learns this relationship and adjusts images to maximize conversions. We derive a theoretical performance guarantee proving the asymptotic optimality of our deep-learning policy. Moreover, A/B testing results indicate that our deep-learning policy substantially increases the platform revenue. Finally, case studies based on a real-life data set also reveal that our policy exhibits good performance even if the functional relationship between images and conversion is misspecified.

4 - Contextual Data-Integrated Newsvendor Solution with Operational Data Analytics (Oda)

Jian Wu, Purdue University, West Lafayette, IN, United States, Qi Feng, J Shanthikumar

We study the data-integrated newsvendor problem in which the random demand depends on a set of covariates. Observing from the solutions analyzed in the existing studies, we identify the equivariant class of operational statistics (i.e., the mapping from the demand and covariate data to the inventory decision) to develop the operational data analytics (ODA) framework for the contextual newsvendor problem. The equivariant property is intuitively appealing, and it is justified by the fact that, regardless of the sample size, no other decision rule can uniformly dominate the optimal operational statistic within the equivariant class. We also demonstrate that nonequivariant solutions can produce unstable empirical performance with limited samples, where equivariant solutions exhibit robustness. When the distribution family of the demand is known but the coefficients of the demand function are unknown, we can directly validate the decision performance of operational statistics within the equivariant class, and derive the uniformly optimal solution. When the distribution family of the demand is unknown, we formulate the data-integration model as a subclass of equivariant operational statistics obtained through adaptively boosting some candidate solution. For decision validation, we project the validation data to the demand for the covariates of interests, and the projection is constructed by utilizing the structure of the candidate solution. We demonstrate the superior small-sample performance of adaptive boosting, and establish the consistency of the boosted operational statistics. Our ODA formulation, building on the inherent characteristics of the contextual newsvendor problem, highlights the importance of understanding structural properties in data-integrated decision making.

SD54

Summit - 446

Causality, Experimentation, and Machine Learning in Digital Platforms

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Ruoxuan Xiong, Emory University, Atlanta, GA, United States

Co-Chair: Wenjia Ba, University of British Columbia, 2053 Main Mall HA576, Vancouver, BC, V6T 1Z2, Canada

1 - Data-Driven Optimization of Experimental Designs

Ruoxuan Xiong, Emory University, DECATUR, GA, United States, Mohsen Bayati

In this paper, we study the optimization of the design of experiments that are run on multiple units for one or multiple time periods. The design problem involves selecting the time periods for the treatment for each unit, with the objective of most precisely estimating the treatment effect. Solving the design problem can be challenging when the expression of the objective function in terms of treatment decision variables is complex and nonexplicit. This can happen when using complex treatment effect estimators or imposing minimal assumptions on the data-generating processes. To address this challenge, we propose a novel zeroth-order Frank-Wolfe algorithm for optimizing experimental designs, that works for general treatment effect estimators and data-generating processes. This algorithm optimizes the design using historical data of the experimental units, which is typically available for applications in digital platforms, clinical trials and other settings. We provide theoretical guarantees for the optimized experimental design from our algorithm. Finally, we show our algorithm yields designs with much lower treatment effect estimation error compared to benchmarks through synthetic experiments on both simulated and real data.

2 - Poisson Limits of Bernoulli Bandits

Wenjia Ba, University of British Columbia, Vancouver, BC, Canada, Lin Fan, Peter Glynn, Michael Harrison

We study Bernoulli bandits in the ``hard luck" parameter regime. That is, the probability of success on any given trial is assumed to be small, regardless of what action the experimenter may choose, and regardless of what exogenous context may be observed. Adopting a general model formulation propounded by Wager and Xu (2021), we prove a limit theorem that justifies approximating a Bernoulli bandit in the hard luck regime by a corresponding Poisson bandit. The latter is a continuous-time treatment selection model in which cumulative reward is driven by one of several possible Poisson processes, depending on the action or treatment selected by the experimenter; each of the Poisson processes has an average success rate that is initially unknown but can be learned through experimentation. Poisson bandits have several advantages over the Bernoulli bandits that they replace or approximate, including both sharper theoretical insights and ease of computation. We introduce policy gradient methods to optimize adjustable Thompson sampling parameters in both discrete-time and continuous-time models. Our experiments indicate that continuous-time policy gradient estimators have lower variance and yield significantly better policies than those from discrete-time policy optimization.

3 - Causal Inference for Large Dimensional Non-Stationary Panels with Two-Way Endogenous Treatment and Latent Confounders

Junting Duan, Stanford University, Stanford, CA, United States, Markus Pelger, Ruoxuan Xiong

This paper studies the imputation and inference for large-dimensional non-stationary panel data with missing observations. We propose a novel method, Within-Transform-PCA (wi-PCA), to estimate an approximate latent factor structure and non-stationary two-way fixed effects under general missing patterns. The missing patterns can depend on the latent factor model and two-way fixed effects. Our method combines a novel within-transformation for the estimation of two-way fixed effects with a PCA on within-transformed data. We provide entry-wise inferential theory for the values imputed with wi-PCA. The key application of wi-PCA is the estimation of counterfactuals on causal panels, where we allow for two-way endogenous treatment effects, time trends and general latent confounders. In an empirical study of the liberalization of marijuana, we show that wi-PCA yields more accurate estimates of treatment effects and more credible economic conclusions compared to its two special cases of conventional difference-in-differences and PCA.

SD55

Summit - 447

Analytics for the Greater Good

Invited Session

Public Sector OR

Chair: Fatemeh Farajzadeh, Worcester Polytechnic Institute, WEBSTER, United States

1 - Proactive Staging of Resources along International Borders for Migratory Crises

Fatemeh Farajzadeh, Worcester Polytechnic Institute, Worcester, MA, United States, Luke Caleb Caddell, Andrew Trapp

Migration crises are complex emergencies that evolve over time and feature high levels of uncertainty concerning human needs. We use stochastic programming to position scarce border resources under demand uncertainty with redistribution recourse decisions. Our costs not only include the economic costs of efficiently operating and securing international borders but also the deprivation experienced as a consequence of accumulated time not accessing critical aid. The unexplored form of interplay between location, resource (re-)allocation, and time introduced in this study enhances strategic and operational decision-making at international borders.

2 - Value Sensitive Design of a Resource Sharing Platform for Nonprofits

Carlos Morel-Figueroa, Rensselaer Polytechnic Institute, Troy, NY, United States, Weixiao Huang, Darren Cole, Yunus Telliel, Sarah Stanlick, Andrew Trapp, Jennifer Pazour

A resource sharing platform for nonprofits, known as SWAP, is a newly developed system piloted by a cohort of Maryland-based nonprofits. Using an integer linear program, SWAP enables participating nonprofits to exchange resources with each other. A common methodology to design technologies that captures the interactions of human values and system design is Value Sensitive Design (VSD). This work uses the VSD methodology to identify the values important to nonprofits and to guide the design of the SWAP system. Our research incorporates the development of different optimization model formulations capturing the nonprofit's identified values into the resource exchange problem, including the implementation of additional SWAP policies and rules of engagement. A series of computational experiments were conducted using empirical data from the SWAP pilot episodes to quantify the proposed designs across the different values.

3 - The Dedicated Docket in US Immigration Courts: An Analysis of Fairness and Efficiency Properties

Daniel Freund, MIT, Cambridge, MA, United States, Wentao Weng

The dedicated docket was introduced by the Biden Administration to reform the immigration system. It creates a separate queue for immigration proceedings where judges are supposed to issue decisions for each case within a target timeframe. Its goals are to improve speed, accuracy, and fairness. Though it meets its first goal, legal advocacy groups report that this comes at the expense of the last. Referring to it as a ``Denial of justice", they find that cases on the dedicated docket routinely fail to access legal representation, and have a much lower asylum grant rate. Against this backdrop, we study the operational implication of the dedicated docket.

We develop a queueing model wherein a policy maker (PM) routes asylees to the regular or the dedicated docket, and sets a delay target for the latter. Constrained by the target, the court allocates its capacity to minimize the average delay. Immigration lawyers schedule their time between dockets to maximize the rate of successful asylum cases. Compared to a single docket, we show that the dedicated-docket system can Pareto-improve the speed and accuracy. However, we also prove that the dedicated docket system satisfies two natural fairness rules only if it is dominated by the single docket.

Our analysis can inform the public discourse. We prove that the lack of fairness is a fundamental design flaw of the new program. Though delay differentiation enables a surprising efficiency gain, this requires unfairness between. Policy makers and legal advocacy groups should be aware of such tradeoffs.

4 - Exposing Human Trafficking Patterns: Leveraging Natural Language Processing for Illicit Classification across US Massage Businesses

Vasuki Garg, North Carolina State University, Raleigh, NC, United States, Maria Mayorga, Osman Ozaltin, Sherrie Caltagirone

The dual nature of massage businesses, firstly as significant outliers among healthcare service facilities due to non-rigorous government scrutiny, and secondly, offering services involving physical touch in distinct privacy, has made them prime candidates for labor and sex trafficking hotspots. This hybrid overlap of human trafficking activities hinders efforts by law enforcement agencies. It also presents a challenge towards the appropriate allocation of valuable resources to identify such illicit businesses effectively. To address this, a unique bifurcated approach to building illicit classification models can be implemented which capitalizes on the business information to understand potential suppliers' aspects and also employs Natural Language Processing techniques to anatomize customer reviews and capture potential consumers' perspectives. With the aim of creating indicators of illicit behavior, these novel models explore the complex consumer–supplier dynamics of human trafficking and highlight patterns of illicit massage businesses to aid classification. These models offer easy deployment for stakeholders to optimally allocate resources. Additionally, using methodologies to identify and incorporate shifts in the indicators when exposed to diverse data offers versatility and adaptability. This allows proficiently analyzing information from different U.S. states, sources, businesses, and domains to identify illicit activities for a larger-scale application of human trafficking network discovery.

5 - A fix-and-optimize matheuristic to solve the location-allocation of vaccination facilities: the case of Jalisco, Mexico.

Marisol Sarai Romero Mancilla, MIT, Boston, MA, United States, Jaime Mora-Vargas, Angel Ruiz

Since the start of the most recent pandemic of COVID-19, around 755 million cases and 7 million deaths have been reported (WHO, 2023). Based on the availability of vaccines to contain the spread of COVID-19, different vaccination plans have been implemented around the world, some being more effective than others, depending on the social, territorial, economic, and political circumstances of each country. In

Mexico and other Latin American countries with similar situations, COVID-19 vaccination has almost exclusively relied on ephemeral mass vaccination facilities together with existing healthcare infrastructure. However, other countries (USA and Canada, for instance) have opted into Pharmacy-based Immunization (PBI), a strategy that consists of using community and/or chain pharmacies as vaccination facilities to provide more accessible immunization services to the population. The aim of this research is to evaluate the feasibility and the expected performance if PBI would have been used in Mexico. To this end, we propose a mathematical formulation to address the location-allocation problem underlying the pharmacy selection and the assignment of individuals to them. However, since commercial solvers are not able to tackle efficiently the resulting formulation for real-sized instances, the formulation is embedded into a heuristic fix-and-optimize scheme allowing to explore the solution space in a more efficient manner. The case of Jalisco, Mexico, is used to illustrate the performance of the proposed approach.

SD56

Summit - 448

Modeling Societal Challenges with OR

Panel Session

Minority Issues Forum

Co-Chair: Himadri Sen Gupta, University of Oklahoma, 202 W Boyd St, Norman, OK, 73019, United States

1 - Moderator Panelist

Himadri Sen Gupta, University of Oklahoma, Norman, OK, United States

2 - Panelist

Andrés González, University of Oklahoma, Norman, OK, United States

3 - Panelist

Karmel Shehadeh, Lehigh University, Bethlehem, PA, United States

4 - Panelist

Senay Solak, University of Massachusetts Amherst, Amherst, MA, United States

5 - Panelist

Robert Curry, University of Arkansas, Fayetteville, AR, United States

SD57

Summit - Terrace Suite 1

Data-driven Modeling for Opioid and Substance-related Public Health Challenges

Invited Session

Health Applications Society

Chair: Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Co-Chair: Zixuan Feng, Penn State, University Park, PA, United States

1 - Measuring Network Dynamics of Drug Overdose Deaths in the United States

Kushagra Tiwari, University of Pittsburgh, Pittsburgh, PA, United States, M. Amin Rahimian, Praveen Kumar, Jeanine M Buchanich, Mark Roberts

The U.S. opioid overdose epidemic has been a significant cause of concern in recent decades. In spite of many studies documenting the role of social networks in overdose use behavior, our current understanding of social network effects in the spatial spread of Opioid Epidemic remains limited. We use Facebook's Social Connectedness Index (SCI) as a proxy for real-life social networks across diverse spatial regions that helps quantify social connectivity across different spatial units. Using Linear Regression, we assess the effect of Opioid Overdose Deaths (OODs) of the alter counties on the ego county. To further validate our findings, we performed a series of robustness checks using cluster-robust standard errors, network autocorrelation, spatial autocorrelation, two-way fixed-effect and two-stage least squares. These checks consistently provide statistically robust evidence of a positive effect of social influence on OODs.

2 - Location and Capacity Planning of Substance use Treatment Facilities for Maximizing Access and Population Health Outcomes

Zixuan Feng, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen, Paul Griffin, Pengyi Shi

The opioid epidemic has devastated communities across the nation, leading to over 500,000 overdose deaths and costing the economy more than $1 trillion over the past two decades. While robust clinical evidence has shown the effectiveness of treatment in reducing mortality and improving recovery, only 1 in 5 individuals who may benefit from medication for opioid use disorders received the treatment. One major barrier to treatment is the insufficient capacity in the current treatment facilities, which makes it difficult for patients who need to travel long distances to access the treatment and thus achieve sustained retention in treatment, particularly for those in rural and underserved areas. However, it remains unclear how the strategic location and capacity planning of treatment facilities can best meet the patient needs to maximize treatment outcomes. In this study, we develop a system dynamics model to formulate the patient flow through the process of waitlist, under treatment, recovery, and relapse for patients from multiple locations. Our model captures the effect of travel distance between patients' locations and the treatment facilities on the treatment retention rate. The policy space includes opening new facilities, expanding current ones, and managing patient assignments and priorities, e.g., delaying admissions based on expected retention rates. We compare these policy combinations by simulating the total operating cost of the treatment facilities and person-months in treatment retention under each policy. We also vary the total budget level and identify the most effective policy that achieves the highest retention outcome under the budget constraint.

3 - Analyzing Patterns of Polysubstance use in Overdose Deaths

Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Drug overdose deaths have continued to increase, with synthetic opioids comprising a more significant proportion. While many studies have assessed the characteristics of individuals who misused and overused a single substance, fewer studies have been conducted to understand patterns of polysubstance use and how the epidemic has evolved over time. This study aims to investigate the patterns of co-occurring substances involved in overdose deaths. We use association analysis and clustering approaches to identify common patterns and evaluate the characteristics of individuals across different groups.

4 - Inventory Modeling for Preventing Opioid Overdose Deaths

RASHID ANZOOM, University of Illinois Urbana-Champaign, Urbana, IL, United States, Rakesh Nagi, Chrysafis Vogiatzis, Hyojung Kang

Federal and state governments in the USA are struggling to contain the opioid epidemic, which has triggered a public health crisis. The increasing cases of overdose fatalities have drawn attention to the distribution of life-saving resources such as Naloxone, prompting efforts to boost their accessibility. This talk explores a data-driven framework to derive resource distribution policies that are cost-efficient, highly responsive, and provide equitable benefits to target regions. Specifically, we focus on defining the appropriate/optimal inventory and transshipment decision parameters under various demand and supply scenarios. We demonstrate the effectiveness of our approach through a simulation model applied to the state of Illinois.

SD58

Summit - Terrace Suite 2

Public Policy and Well-Being

Invited Session

Health Applications Society

Chair: Anqi Wu, Florida International University, Miami, FL, United States

1 - The Development of Precision Medicines: An Economic Impediment OR a New Model of Drug Innovation?

Anant Mishra, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Jingwen Yang

In recent years, the emergence of precision medicines has transformed the practice of medicine. Yet, the strategic implications of such innovations on market dynamics remain unclear. Does precision medicine curtail the market size by focusing on niche patient populations or does it enhance overall market performance due to heightened drug efficacy? Leveraging FDA approvals of precision indications for identification, we uncover that, contrary to conventional believe of market contraction, precision medicines lead to a significant market expansion. Specifically, approvals of precision indications are associated with increases in drug sales by 96% and in revenues by 90%. The presence of companion diagnostics, used exclusively for patient identification, amplifies these effects. The research findings suggest that drug markets respond to precision medicine innovation despite the concern of such innovation targeting narrower market segments. Business model challenges tied to this innovation can be mitigated by the presence of companion diagnostic devices. In essence, precision medicine innovation signifies a shift from the conventional blockbuster drug business approach, offering a promising alternative strategy for drug innovation.

2 - Opioid Control Policies Can Also Reduce Domestic Violence

Minglu Sun, RTI International, Tampa, FL, United States, Andrei Barbos

Opioid abuse is an issue of serious concern in the United States, and it has been the focus of a multitude of state and federal level policies. Such policies can raise cost vs. benefit considerations, which besides direct effects, must also account for potential second-order unintended consequences. We provide evidence for an important positive spillover of effective opioid control policies by showing that they can also reduce domestic violence. To this aim, we exploit the staggered implementation of the Mandatory Access Prescription Drug Monitoring Programs, which required health care providers to consult an electronic database before prescribing and/or dispersing controlled substances. These programs have been shown to be effective at reducing the utilization of opioids. We find that they also decreased the instances of intimate partner assaults, driven primarily by a decrease in simple assaults, although there is evidence that they also decreased the incidence of the other types of assaults. The results are robust under multiple specifications.

3 - The Difficulty of Achieving Gender Equality: Evidence from US Universities

Tiancheng Zhao, Saint Louis University, St. Louis, MO, United States, Sridhar Seshadri, Zur Shapira

Despite progress in reducing gender inequality, disparities continue in the academic labor market, particularly regarding vertical gender segregation. To investigate the feasibility of achieving gender equality at the highest academic levels in US universities, we develop a mathematical model representing promotion and hiring. Our findings, using data from 2001 to 2019, show that certain universities tend to have more favorable promotion rates for male professors than for females—for example, 15.95% higher at the assistant professor level. We discuss university-level characteristics that influence inequality. Numerical simulations reveal that, even with favorable treatment, existing gender imbalances impede progress at higher levels, with 34.9 years on average to achieve equal representation at the full professor level. This research provides valuable insights into challenges associated with promoting gender equality in academia.

SD59

Summit - Ballroom 1

APS Distinguished Speakers: Mengdi Wang and Minshuo Chen

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 -

2 - A Probabilistic Foundation And New Opportunities of Diffusion Models for Generative AI

Minshuo Chen, Northwestern University, Plainsboro Township, NJ, United States, Mengdi Wang

Diffusion models, a powerful and universal generative AI technology, have achieved tremendous success with multi-modality data from engineering, operations research, computer science, etc.

In this tutorial, we aim to unveil and capitalize diffusion models. Firstly, we review the applications and design philosophy of diffusion models from a probabilistic point of view. Secondly, we overview the existing theories connecting the probabilistic nature of diffusion models to their statistical properties and sampling capabilities. Thirdly, we introduce brand new opportunities embodied by diffusion models, where we highlight one example of diffusion models for generative optimization.

SD60

Summit - Ballroom 2

Strategic Supply Chain Network Optimization at Walmart

Invited Session

The Practice Section of INFORMS

Chair: Xinyue Peng, Walmart, Sunnyvale, CA, United States

1 - Large Scale Optimization Models for Evolving Walmart Next-Gen Ambient Network

Xinghua Wang, Walmart, Dallas, TX, United States, Mehmet Aydemir, Pengcheng Zhang, Sai Rajesh Mahabhashyam

Connecting vendors and more than 4600 stores in the US, Walmart’s Regional Distribution Centers (RDCs) play a critical role in its supply chain network. Where to set up these facilities, how much capacity each should have, and which stores should be supported by each RDC have significant impact on the retailer’s overall supply chain cost, efficiency and resilience. To further improve operational efficiency and lower cost, we are planning to upgrade existing RDCs to meet future demand. Key questions that need to be answered for such transformation include: 1) what’s the best timeline of the upgrades; 2) do any of the RDCs need to be supported by other RDCs during the upgrade process; 3) what are the best store-RDC alignments during and after the transformation; 4) how to better integrate with other supply chain facilities with the extra capacity that becomes available after the upgrades.

To answer these questions and ensure a smooth transition, we build a large-scale, multi-period dynamic Mixed Integer Programming model to minimize overall transportation and handling costs while observing various realistic operational constraints. To solve the model efficiently, we decompose the problem into two stages. The first stage involves finding the optimal locations and configurations for the end state. The second stage then finds the optimal strategy to transform the current RDC network to the next-gen network. In this talk, we will introduce the business context, present the mathematical model formulation, and illustrate the solution approach.

2 - Multi-commodity Network Flow Optimization for Walmart Middle Mile E-commerce Network

Anand Seshadri, Walmart E-commerce, Dallas, TX, United States, Aida Khayatian, Sai Mahabhashyam

For the Walmart e-commerce supply chain strategy team to make decisions regarding the optimal evolution of the Walmart network in the future, a robust software application is necessary that allows for rapid prototyping, data setup at scale, extensible experimentation, and detailed feedback. The purpose of this paper is to describe state of the art solution approach utilized by to formulate the middle mile strategy. The output of the application describes various product flows, paths utilized, and the choice of cross-dock nodes given the potential demand and allowable cross-dock sites which include sort centers and delivery stations. The solution also must include various business constraints such as delivery time, minimum and maximum volumes at the cross-dock nodes, and maximum number of intermediate nodes along a path. The solution also must satisfy multiple business objectives such as minimum cost, and maximum last-mile delivery volume. The solution presented solves the problem inside a modified multi commodity network flow framework and enables fast optimal solutions to be obtained quickly (within an hour) for large (~3,000 nodes and 400,000 arcs) network. The model also is executed sequentially across multiple years and ensures network consistency across the time-horizon under consideration.

3 - Optimization Models for Reducing Empty Miles in Walmart Transportation Network

Rahul Swamy, Walmart Centroid, Dallas, TX, United States

Walmart operates one of the largest transportation fleets in the United States, facilitating the movement of loads across an extensive network that includes vendors, distribution centers, consolidation centers, and retail outlets. Despite the expansive nature of this network, a significant portion of freight from vendors is currently transported by third-party carriers. Walmart's key strategic goal is to internalize a portion of these loads to enhance overall network efficiency and achieve cost savings. This talk introduces a multi-objective optimization model designed to identify and prioritize vendors whose freight can be internalized into Walmart's fleet operations. The proposed assignment model ranks vendors while optimizing several metrics, such as minimizing network inefficiencies (e.g., reducing empty miles), maximizing alignment with the existing fleet network, and enhancing temporal consistency in shipping lanes. The model is solved using the epsilon-constraint method. By internalizing these freight, Walmart's network efficiency is expected to improve, thereby reducing carbon emissions and increasing cost savings.

4 - Densifying Walmart Last Mile Delivery via Trip Planning Optimization

Lingxun Kong, Walmart Centroid, Sunnyvale, CA, United States, Yao Luo

At Walmart, we manage deliveries under various service level agreements (SLA). For scheduled deliveries, orders are delivered within a time window, whereas for unscheduled deliveries, the deliveries can be at any point throughout the day. To densify our last-mile delivery system, we are commingling both scheduled and unscheduled orders in the same trip. If we are given all scheduled and unscheduled orders for the day, we can plan commingled trips by solving a capacitated vehicle routing problem with time windows (CVRPTW). In operations, however, orders are placed continuously throughout the day, necessitating multiple trip planning sessions. The challenge lies in intelligently selecting unscheduled orders for delivery during the slots in the morning, without prior information about the orders that will be placed later. To tackle this, we amalgamate existing orders with a set of predicted synthetic orders, and solve the CVRPTW for the entire day, executing deliveries only for the current slot. This process is repeated in subsequent trip planning, using updated actual and synthetic orders for the remaining horizon. This method allows us to plan cost-effective trips for the receding horizon based on our best future estimates. Implementing this framework could result in notable cost savings and SLA enhancements for Walmart's last-mile ecosystem.

5 - Model Approach for Solving Large Scale Affinity Aware Inventory Placements at Walmart

Yaqing Wu, Walmart Centroid, Dallas, TX, United States, Bharath Rangarajan

A large number of orders are dispatched from tens of fulfillment centers at Walmart every year to satisfy the customer orders placed online. Only a selected set of items could be placed in a fulfillment center considering physical or operational capacity constraints. Order splits happen if items in an order are fulfilled from multiple fulfillment centers. Thus, it is desired to place items frequently purchased together (that is, high basket-affinity items) in the same fulfillment centers. The problem becomes complex as Walmart offers large assortments for customers to consider, it is challenging to decide the placement of items across the fulfillment network. We propose an end-to-end, scalable optimization framework for this placement problem. We run experiments to show significant performance improvements in end-to-end order fulfillment cost.

SD61

Summit - Ballroom 3

Solo-preneur Consulting

Panel Session

The Practice Section of INFORMS

Co-Chair: William Christian, U.S. Dept of Defense, Severn, MD, United States

1 - Moderator Panelist

2 - Panelist

Carrie Beam, UC Davis, Sacramento, CA, United States

3 - Panelist

Lincoln Chandler, Chandler Decision Services, Chicago, IL, United States

4 - Panelist

Zohar Strinka, Analytics Strategies LLC, Thornton, CO, United States

5 - Panelist

Erick Wikum, Wikalytics, LLC, Maineville, OH, United States

SD62

Summit - Signature Room

Experimental Design For Causal Inference Through An Optimization Lens

Invited Session

TutORial

Chair: Harish Krishnan, University of British Columbia, Vancouver, BC, Canada

1 - Experimental Design For Causal Inference Through An Optimization Lens

Jinglong Zhao, Boston University, Boston, MA, United States

The study of experimental design offers tremendous benefits for answering causal ques tions across a wide range of applications, including agricultural experiments, clinical trials, industrial experiments, social experiments, and digital experiments. While valu able in such applications, the costs of experiments often drive experimenters to seek more efficient designs. Recently, experimenters have started to examine such efficiency questions from an optimization perspective, as experimental design problems are fun damentally decision-making problems. This perspective offers a lot of flexibility in leveraging various existing optimization tools to study experimental design problems.

thus aims to examine the foundations of experimental design problems in the context of causal inference as viewed through an optimization lens.

SD63

Regency - 601

Human-AI Interaction in Business

Invited Session

Information Systems

Chair: Qinglai He, University of Wisconsin - Madison, Madison, WI, United States

1 - On the Double-Edged Effects of Peer-to-Peer Payment Systems in Consumer Banking

Ao Huang, University of Miami, Coral Gables, FL, United States, Ni Huang, Yili Hong

As a fast-growing financial technology, peer-to-peer (P2P) payment systems like Zelle offer consumers a convenient way to transfer money. Anecdotal evidence suggests the convenience of short-term liquidity management may come at the unintended cost of increased fraud. This study empirically investigates the effects of Zelle availability, focusing on bank consumers' temporary financial hardship and fraud complaints. Leveraging the staggered partnership with the Zelle payment network across U.S. banks from 2017-2021 and consumer complaint data from the Consumer Financial Protection Bureau (CFPB), we employ difference-in-differences (DID) analyses and counterfactual estimations to identify the said effects. The results show that Zelle availability led to a 14.1% decrease in complaints related to the struggle with student loan payments but a 22.9% increase in fraud complaints. Larger banks and counties with larger African American populations experienced more fraud complaints post-Zelle implementation. Zelle availability appears to present a double-edged effect - while it may help alleviate temporary financial hardship, it also increases fraud risk, disproportionately affecting socioeconomically disadvantaged groups. Our findings highlight the need for policymakers and financial institutions to enhance consumer protection and risk management efforts when implementing P2P payment technology.

2 - The Potential of Large Language Models in Revolutionizing Charitable Giving

Teng Ye, University of Minnesota, MINNEAPOLIS, MN, United States, Jingnan Zheng, Junhui Jin, Jingyi Qiu, Wei Ai, Qiaozhu Mei

Online charitable giving has provided essential alternative funding source, particularly for small businesses that often face high financial vulnerability and limited access to traditional financial support compared to larger corporations. However, as small businesses increasingly turn to online crowdfunding platforms, over 40% of these campaigns fail to raise any funds, especially those operating in low socio-economic regions. We leverage the latest advancements in AI technology to identify and strategically optimize key factors that influence the success of crowdfunding campaigns. By analyzing over 10,000 small business charitable-giving campaigns in the U.S., our best-performing machine learning model accurately predicts the fundraising outcomes of 81% of campaigns. Interpreting the machine learning model enables us to offer actionable suggestions for improving the textual description before launching a campaign. By enhancing just three aspects of its description using a large language model, a campaign becomes more preferable to 83% human evaluators. This increases the likelihood of securing financial support for originally unfunded campaigns by 13.3% and significantly improves the equity of fundraising results. Our research uncovers effective strategies for crafting descriptions for small business fundraising campaigns and opens up a new realm in integrating large language models into online charitable giving methodologies.

3 - A Simple Solution to Improving Human Supervision of Algorithms: Evidence from Smart Vending

Brian Han, Gies College of Business, UIUC, Champaign, IL, United States, Minda Zhao, Xin Chen, Tao Zhu

Humans nowadays inevitably face the human-AI interaction problem of whether to follow algorithms' suggestions. Although studies indicate human overrides degrade performance, humans sometimes possess private information that increases the profit. This paper proposes a simple solution to improve human supervision by allowing human overrides with constraints. Compared to the binary choice to either fully permit or prohibit human overrides, constraints incentivize humans to make selections more judiciously and override only with high confidence to improve performance. We partnered with FengE ZuShi (a large smart-vending company in China) and conducted a field experiment to test our solution in the context of inventory replenishment. We find strong evidence supporting our solution that constrained override increases sales by 1.3%. Besides, our method is a low-cost approach in practice and scales easily to a broader class of business operations as imposing override constraints is a general solution.

4 - Traffic Allocation Optimization on Video Sharing Platforms: Balancing Viewer Engagement and Small Creator Development

Qinlu Hu, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Ni Huang, Renyu Zhang

Online short-form video platforms like TikTok, Triller, and Snapchat Spotlight, driven by algorithms that enhance viewer engagement, often favor established creators over emerging, small creators. This study focuses on small creators, seeking solutions to balance viewer engagement and creator development via viewing traffic allocation. In collaboration with a leading online short-form video sharing platform, we report two field experiments. The partner platform conducted a viewer-side experiment to limit the viewers' access to the small creators' videos, and our analysis shows that reducing small creators' content generally boosted viewer engagement. However, effects varied by viewer activeness -- the treatment increased video watching time for less active viewers but decreased it for active viewers. The partner platform also conducted a creator-side experiment, allocating extra viewer traffic to small creators. This led to significant increases in their content production quantity, efforts and new follower count, with insignificant changes in content quality. Notably, these effects were more pronounced among the more experienced and popular small creators. Combining all insights, we propose a novel traffic allocation policy: boosting the content visibility of small creators, especially those who are relatively more popular, to active viewers, while limiting it for less active viewers in the recommender system of the platform. This strategy enhances engagement for both viewer types and improves small creators' productivity and growth. This approach has the potential to create a tripartite win-win-win outcome, indicative of a Pareto improvement that benefits the creators, viewers, and the platform. We discuss the theoretical and practical implications.

5 - The Impact of Bifurcation on Platform Outcomes in a Q&A Community

Xiaomeng Chen, University of Pittsburgh, Pittsburgh, PA, United States

This paper studies platform bifurcation, the process in which a subgroup of users from an original platform launches an independent spin-off platform. We identify the effects of bifurcation using a difference-in-differences approach that exploits the introduction of spin-off platforms in an online platform incubator. We find that bifurcation leads to a strong overall increase in contributions. While contributions to the home platform decline, the two bifurcated platforms generate more combined user contributions and attract more new users compared to a single united platform. We further explore how interconnectivity and platform differentiation affect users' choice of platform. Our evidence indicates that users are less likely to migrate from an incumbent platform to a new, specialized platform when interconnectivity is strong but more likely to migrate when the specialized platform enables more differentiation. This paper is the first to empirically analyze the strategic implications of new platform entry at scale and to document the moderating role of interconnectivity and platform differentiation.

SD64

Regency - 602

Data-Driven Decision Making: Harnessing Multimedia, Social Media, Big Data, and Machine Learning for Business Excellence

Invited Session

Social Media Analytics

Chair: Cuibing Wu, Framingham State University, Framingham, MA, United States

1 - Impact of Verbal and Nonverbal Cues in CEO Interview Videos on Firm Risk

Cuibing Wu, University of Massachusetts Lowell, Lowell, MA, United States, Julie Zhang, Remi Daviet

This study investigates the impact of verbal and nonverbal cues in CEO interview videos on stock volatility, analyzing a dataset of 2,404 CEO interviews from S&P 500 companies from 2010 to 2022. Using advanced machine learning techniques, the research quantifies verbal, vocal, and visual cues to examine their relationship with stock volatility. Empirical findings indicate a strong correlation between the elements of CEO interviews and stock volatility. In the experimental phase, GANs are used to generate varying emotional expressions in CEO videos, and participants are shown these videos. The results reveal that videos with happy CEO expressions lead to expectations of increased stock value, while those with sad or surprised expressions predict a decrease. This comprehensive study highlights the significant influence of CEO verbal and nonverbal cues on stock volatility.

2 - Taste or Terminate? Deciphering Factors of Restaurant Failure Through Customer Reviews

Jie LI, University of Massachusetts Lowell, Lowell, MA, United States, Asil Oztekin

This study examines the dynamics of social interactions and the impact of online reviews on Yelp, utilizing the Yelp Academic dataset to conduct a comprehensive analysis. We focus on extracting text features from consumer reviews, integrating a unique method to calculate an aggregated review score that assigns weights to individual reviews based on specific criteria. The study further explores the predictive power of restaurant features and textual data in forecasting restaurant closures. We compare the performance of several machine learning models, including ZeroR, Multi-layer Perceptron (MLP), and Logistic Regression, to assess their effectiveness in this context. Our predictive model employed is a custom-built neural network featuring ReLU activation functions within its hidden layers.

3 - When Carryover Meets Spillover: Incrementality Measurement in Catalog and Beyond

Zhengrong Gu, Boston University, Boston, MA, United States, Garrett Johnson, Ziwei Liao, Jerry Chen

Marketing channel carryover effects occur when the channel has the persistent effect of advertising after exposure. In particular, channels such as direct mail catalogs can exhibit carryover effects long after being sent. Moreover, channel carryover effects may interact with channel spillover effects: i.e., the combined effect of two marketing channels (e.g., email and direct mail catalogs) is greater than or less than the sum of each channel individually. Measuring complementarity or cannibalization between channels helps marketers understand and optimize their spend holistically. This paper addresses both spillover and carryover effects in a multi-channel marketing context at a global e-commerce firm with millions of customers. We use large-scale lift studies to measure the effectiveness of direct mail catalog channel with different numbers of catalogs and test durations. We find long-term carryover effects that persist for months after the mailing, with more substantial effects noted for groups receiving multiple seasonal catalogs. We then examine how subsequent exposures via routine email marketing campaigns interact with the experimentally measured effect of the direct mail catalog channel. This research shows the extent and nature of channel spillover effects and reveals how carryover effects interact with spillover effects, enhancing the understanding of channel interactions.

4 - Cracking the Ca(u)ses: Unraveling Urban Crime Dynamics with Agent-Based Modeling

Ann-Kathrin Meyer, University of Münster, Münster, Germany, Konstantina Valogianni, Balaji Padmanabhan, Tobias Brandt

Understanding the dynamics of urban crime is a vital aspect of contemporary societal management, enabling proactive measures for law enforcement and policy makers. However, traditional crime prediction models often fall short in capturing the intricate causal mechanisms driving criminal activities, relying heavily on statistical correlations.

Agent-Based Modeling (ABM) presents a unique approach for deriving causal frameworks and we demonstrate its potential in capturing complex relationships inherent in criminal behavior. Specifically, we employ ABM to construct causal models to explain crime dynamics in Los Angeles using, among others, georeferenced social media data as a measure of social activity. The design and implementation of Causal ABMs and leveraging genetic algorithms for their estimation enables the extraction of multiple plausible causal sets from empirical data.

Through the understanding of crime causes, we aim to pave the way for the development of targeted preventive strategies. Furthermore, this approach empowers researchers to develop robust causal ABMs that offer a deeper understanding of crime dynamics.

5 - Navigating the Storm: The Impact of Social Media Crisis Response Strategies on Stock Performance

ziqian song, The university of scranton, Dalton, PA, United States

In the contemporary business landscape, social media platforms have emerged as critical arenas for crisis communication, significantly influencing stakeholders' perceptions and market performance. This study explore the interplay between firms' crisis response strategies deployed on social media and their consequential impact on stock performance. Leveraging the Situational Crisis Communication Theory (SCCT) framework, we examine the effectiveness of various crisis response tactics employed by firms across social media channels. Through empirical analyses, we evaluate firm crisis response behaviors on social media and their subsequent influence on stock market performance. Our investigation analyze the extent to which firms' strategic communication endeavors during crises shape investor perceptions and subsequently impact market outcomes. Our findings not only illuminate the efficacy of social media crisis response strategies in mitigating financial influence but also provide actionable insights for firms endeavoring to adeptly navigate social media crises while upholding shareholder interests. This research contributes valuable knowledge to firm crisis communication using social media and offers practical implications for firms seeking to bolster their crisis management capabilities in the digital age.

SD65

Regency - 603

Reimagining Assessments, Unleashing Creativity, and Grading with GenAI

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Graeme Warren, Johns Hopkins University, Rockville, United States

1 - Moderator Panelist

Graeme Warren, Johns Hopkins University, Washington, DC, United States

2 - Panelist

Geoffrey Pofahl, Arizona State University, Mesa, AZ, United States

3 - Panelist

4 - Panelist

Lakshmi Iyer, UNC Greensboro, Greensboro, NC, United States

SD66

Regency - 604

Stochastic Approximation (SA) and Stochastic Gradient Descent (SGD) for Machine Learning

Invited Session

Artificial Intelligence

Chair: Siva Theja Maguluri, ISyE Georgia Tech, Atlanta, GA, United States

1 - Two-time-Scale Linear Stochastic Approximation: Tight Finite Time Bounds Under Markovian Noise

Shaan Ul Haque, Georgia Institute of Technology, Atlanta, GA, United States, sajad Khodadadian, Siva Theja Maguluri

Stochastic approximation (SA) is an iterative algorithm to find the fixed point of an operator given noisy samples of this operator. SA appears in many areas such as optimization and Reinforcement Learning (RL). When implemented in practice, the noise that appears in the update of RL algorithms is naturally Markovian. Furthermore, in some settings, such as gradient temporal difference (GTD), SA is employed in a two-time-scale manner. The mix of Markovian noise along with the two-time-scale structure results in an algorithm which is complex to analyze theoretically. In this talk, we characterize a tight convergence bound for the iterations of linear two-time-scale SA with Markovian noise. Our results show the convergence behavior of this algorithm given various choices of step sizes. Applying our result to the well-known temporal difference with gradient correction (TDC) algorithm, we show the first $O(1/\epsilon)$ sample complexity for the convergence of this algorithm, outperforming all the previous work. Similarly, our results can be applied to establish the convergence behavior of a variety of RL algorithms, such as TD-learning with Polyak-Ruppert averaging, GTD, and GTD2.

2 - Federated Offline Reinforcement Learning: Collaborative Single-Policy Coverage Suffices

Jiin Woo, Carnegie Mellon University, Pittsburgh, PA, United States, Laixi Shi, Gauri Joshi, Yuejie Chi

Offline reinforcement learning (RL), which seeks to learn an optimal policy using offline data, has garnered significant interest due to its potential in critical applications where online data collection is infeasible or expensive. This work explores the benefit of federated learning for offline RL, aiming at collaboratively leveraging offline datasets at multiple agents. Focusing on finite-horizon episodic tabular Markov decision processes (MDPs), we design FedLCB-Q, a variant of the popular model-free Q-learning algorithm tailored for federated offline RL. FedLCB-Q updates local Q-functions at agents with novel learning rate schedules and aggregates them at a central server using importance averaging and a carefully designed pessimistic penalty term. Our sample complexity analysis reveals that, with appropriately chosen parameters and synchronization schedules, FedLCB-Q achieves linear speedup in terms of the number of agents without requiring high-quality datasets at individual agents, as long as the local datasets collectively cover the state-action space visited by the optimal policy, highlighting the power of collaboration in the federated setting. In fact, the sample complexity almost matches that of the single-agent counterpart, as if all the data are stored at a central location, up to polynomial factors of the horizon length. Furthermore, FedLCB-Q is communication-efficient, where the number of communication rounds is only linear with respect to the horizon length up to logarithmic factors.

3 - Order-Optimal Convergence Rates with Adaptive Sgd

Matthew Faw, Georgia Institute of Technology, Atlanta, GA, United States

We study convergence rates of AdaGrad-Norm as an exemplar of adaptive stochastic gradient descent (SGD) methods, which adjust their step sizes based on observed stochastic gradients, for minimizing non-convex, smooth objectives. Despite their popularity, the analysis of adaptive SGD lags behind that of non-adaptive methods in this setting. Specifically, all prior studies rely on some subset of the following assumptions which are unnecessary in the analysis of fixed step-size SGD: (i) uniformly bounded gradient norms, (ii) uniformly bounded stochastic gradient variance (or even noise support), (iii) conditional independence between the step size and stochastic gradient. In this work, we show that AdaGrad-Norm exhibits an order optimal (up to polylogarithmic factors) convergence rate scaling inversely with the square-root of T after T iterations under the same standard assumptions as a well-tuned fixed step-size SGD (unbounded gradient norms and affine noise variance scaling). We also go beyond the standard smoothness assumption, proving similar results for (L_0, L_1)-smooth (non-convex) functions (where the Hessian can scale affinely with the gradient, e.g. such as the exponential function). This is the first rate of convergence for any algorithm in this smoothness and noise regime, even for the special case of uniformly-bounded variance. This talk is based on joint works with Isidoros Tziotis, Litu Rout, Constantine Caramanis, Aryan Mokhtari, Rachel Ward, and Sanjay Shakkottai. References: https://arxiv.org/abs/2202.05791 (COLT 2022) and https://arxiv.org/abs/2302.06570 (COLT 2023).

4 - Rates of Convergence in the Central Limit Theorem for Markov Chains, with An Application to Td Learning

R Srikant, University of Illinois at Urbana Champaign, Urbana, IL, United States

We prove a non-asymptotic central limit theorem for vector-valued martingale differences using Stein's method, and use Poisson's equation to extend the result to functions of Markov Chains. We then show that these results can be applied to establish a non-asymptotic central limit theorem for Temporal Difference (TD) learning with averaging.

SD67

Regency - 605

Challenges and Innovations in Sustainable Supply Chains

Invited Session

MSOM: Sustainable Operations

Chair: Xavier Warnes, Indiana University, Kelley School of Business, Bloomington, IN, United States

Co-Chair: Can Zhang, Duke University, Durham, NC, United States

1 -

Tip Your Farmer? Implications of Tipping in Agriculture on Sustainability and Financial Inclusion

Saed Alizamir, University of Virginia, Charlottesville, VA, United States, Basak Kalkanci

An emerging financial innovation enabled by technological advancements in agricultural supply chains is the capability to "tip the farmers." This innovation empowers socially-conscious customers to identify the individual farmers of their sustainably-sourced products and reward these farmers by sending them direct payments, or tips, through mobile apps. We examine the implications of this new capability on farmers' and consumers' welfare and agricultural firm profits. We characterize the equilibrium of the game with and without tipping and show how different types of equilibria emerge depending on the model parameters. We investigate the implications of tipping on various stakeholders such as farmers, customers, and the agricultural firm and identify the conditions under which they may be better off or worse off. Furthermore, we find that even when farmers benefit from tipping in expectation, this financial innovation may create disparity and exacerbate inequity in the farmer population, which is undesirable from a social responsibility standpoint. Firms must exercise caution in implementing the tipping capability as farmers' expected and actual income and consumer welfare may reduce in the presence of tipping. In contrast, if tipping is implemented under the right conditions, it can create a triple win for agricultural firms, farmers, and customers.

2 - Empowering OR Exploiting? the Implications of Direct Market Access for Improving Smallholder Farmers&Rsquo; Welfare

Xi Lin, The Paul Merage School of Business, UC Irvine, Irvine, CA, United States, Luyi Gui, Yixin Lu

Poor market access is a major hurdle to poverty reduction for smallholder farmers. Traditional market access strategies that connect farmers with wholesale intermediaries are criticized for exposing farmers to exploitation. While many believe that this issue can be resolved by direct market access under which farmers sell directly to consumers, practice indicates that farmers often need to work with service intermediaries to facilitate such direct sales and continue to suffer from exploitation. Hence, it remains unclear whether and under what circumstances direct market access can benefit smallholder farmers. In this paper, we address these questions through a comparative study of two prevalent strategies: contract farming, which provides farmers with assured sales of their output to a buying firm at a pre-determined wholesale price, and rural livestreaming, whereby farmers sell to consumers in a live broadcast run by a media company that charges a percentage commission fee. We show that compared to contract farming, the direct market access enabled by livestreaming can mitigate exploitation and improve farmers' income if they plant niche crops or incur low planting costs. Otherwise, direct market access may aggravate exploitation and hurt farmers' income. Furthermore, direct market access can become more effective in improving farmers' income under yield uncertainty or if farmers' bargaining power improves, but may cause agricultural subsidies to backfire and reduce farmers' income. Our results indicate that policy makers should account for the operational characteristics of local crops and other poverty reduction policies when choosing between direct versus indirect market access strategies.

3 - Urban Mining and Material Scarcity in Clean Energy Supply Chains

Clara Carrera, INSEAD, Fontainebleau, France, Atalay Atasu, Luk Van Wassenhove, Serasu Duran

The low-carbon economy is a materials economy. Clean energy technologies – essential for the reduction of carbon emissions – require large amounts of critical raw materials such as cobalt for energy storage, rare earth minerals for wind turbines, and silver for solar panels. The supply of these critical materials cannot keep up with skyrocketing demand, which may be a roadblock against the ongoing clean energy transition. In this paper we study two practical approaches that can sustain the clean energy industry's momentum: (i) Material Reduction, i.e., changes to production technology to reduce the critical materials used; (ii) Urban Mining, i.e., recovering and recycling critical materials from end-of-life clean energy technology products. We show that the effectiveness of these two approaches depends on the levels of material scarcity and systemic leakage in circular infrastructures. A long-term focus on clean energy production favors Urban Mining (Material Reduction) when material scarcity is high (low). Furthermore, there is an optimal level of circularity in the design of urban mining systems that maximizes clean energy generation in the long run. Meanwhile, an urban mining strategy is the most likely to maintain the profitability of the clean energy industry. Policy support to moderate the capital costs and systemic leakage in circularity infrastructures can motivate the clean energy industry to invest in urban mining.

SD68

Regency - 606

Nicholson Student Paper Competition I

Award Session

Nicholson Student Paper Competition

Chair: Siqian Shen, U of Michigan/NSF, Ann Arbor/Alexandria, 48109, United States

Co-Chair: Andrew Daw, USC Marshall Data Sciences and Operations, Los Angeles, CA, United States

1 - Sensitivity analysis for mixed binary quadratic programming

Jingye Xu, Georgia Institute of Technology, Atlanta, GA, United States

We consider sensitivity analysis for Mixed Binary Quadratic Programs (MBQPs) with respect to changing right-hand-sides. We leverage Burer's completely-positive (CPP) reformulation of MBQPs. Its dual is an instance of co-positive programming (COP) and can be used to get sensitivity bounds. We give a constructive proof that strong duality between the CPP and COP problems holds iff the feasible region is bounded, or the objective function is convex. We provide a method for finding good nearly optimal dual solutions and present preliminary computational results on sensitivity analysis for MBQPs.

2 - Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods

Shuvomoy Das Gupta, Columbia IEOR, NYC, NY, United States

We present BnB-PEP, the first unified methodology for constructing the optimal first-order methods for both convex and nonconvex optimization. BnB-PEP poses the problem of constructing the optimal method as a nonconvex but practically tractable quadratically constrained quadratic optimization problem and solves it to certifiable global optimality using a customized branch-and-bound algorithm. We apply BnB-PEP to various practically relevant setups to construct optimal methods that surpass previous state-of-the-art results and also systematically generate analytical convergence proofs.

3 - The (Surprising) Sample Optimality of Greedy Procedures for Large-Scale Ranking and Selection

Zaile Li, Fudan University, Shanghai, China, People's Republic of

Ranking and selection (R&S) aims to identify the best alternative from a finite set by mean performance. Ideal large-scale R&S procedures should be sample optimal, i.e., the total sample size for a nonzero probability of correct selection (PCS) grows linearly with the number of alternatives. Surprisingly, we find that the naïve greedy procedure appears sample optimal. We develop a boundary-crossing perspective to understand the discovery, prove the sample optimality, and further analyze the probability of good selection. Our discovery may suggest a new mindset for designing R&S procedures.

4 - Differential Privacy via Distributionally Robust Optimization

Aras Selvi, Imperial College Business School, London, United Kingdom

Differential privacy, the de facto standard for privately sharing statistics about datasets, inherently involves a trade-off between privacy and accuracy of the released information. We develop a class of mechanisms that optimally balance these conflicting objectives. To this end, we formulate the mechanism design problem as an infinite-dimensional distributionally robust optimization problem and show that the problem affords a strong dual, and we exploit this duality to develop converging hierarchies of finite-dimensional upper and lower bounding problems.

SD69

Regency - 607

AI and AI

Invited Session

eBusiness

Chair: Yumei He, Tulane University, New Orleans, LA, 70119, United States

1 - Clean up the Act: Impact of Privacy Regulation on Live Streaming Influencers

Yuan Yuan, Carnegie Mellon University, Pittsburgh, PA, United States, Nikhil Malik, Wen Wang, Kannan Srinivasan

In the digital age, live-streaming influencers on platforms like TikTok and Douyu.com have seen a surge in popularity. These influencers often strategically leverage a portion of their "privacy assets", such as by revealing private details, to garner rewards like engagement and intimacy. How- ever, this strategy has led to significant privacy concerns and safety issues, prompting policy regulations in recent years. This study empirically investigates such phenomena and quantifies the impact of privacy regulations on influencers. Our analysis, using observational data from Douyu.com and multimodal video analytics, shows that self-disclosure is associated with higher audience engagement for influencers. Leveraging a natural policy shock, we find that regulatory policies on such behavior have led to a notable decrease in audience engagement, with heterogeneous effects across influencers creating different styles of content, suggesting widespread economic losses due to these regulations. Interestingly, midst the general negative effects on the platform, some influencers have proactively adapted by switching to different styles of content creation, resulting in in- creased engagement. We leverage a game theoretical model to explore the underlying mechanisms and examine the reasons behind influencers' decisions to change content categories in response to policy shocks. This study provides vital insights into the dynamics of the digital economy, especially in the live-streaming domain. It highlights the significant implications for future policy development, platform management, and the evolution of content creation strategies, balancing live-streaming costs with the pursuit of gains on online platforms.

2 - She? the Role of Perceived Agent Gender in Social Media Customer Service

Junyuan Ke, University of Rochester, Rochester, NY, United States, Yang Gao, Huaxia Rui, Shujing Sun

Understanding the effect of gender in business interactions is of significant importance to companies and society. Inspired by practical concerns about how gender affects customer support interactions, the present study investigated the role of perceived agent gender in customer behavior using a unique dataset consisting of all public customer service interactions handled on Southwest Airlines' Twitter platform from March 2018 to September 2019. We inferred agent gender based on the first names provided by agents when responding to customers. In addition, we measured customer behavior using the following outcomes: whether a customer decided to continue the service conversation upon receiving an agent's initial response as well as the emotional reactions in their second tweet if the customer chose to continue the interaction. Our identification was based on the backdoor criterion and hinged on the assumption that customers are assigned to the next available agent, independent of agent gender. The findings revealed that customers were more likely to continue interactions with female agents than with male agents. Customers' emotional reaction to perceived agent gender is more nuanced. We find that customers' messages were more negative in valence but less intense in arousal with the former group than with the latter. These results offer valuable insights both to practitioners and to academics regarding gender stereotypes and bias in the service industry.

3 - Empathic Algorithm Delegation: Fostering Optimal Human-AI Collaboration in Decision-Making

Thomas Ware, Arizona State University, Tempe, AZ, United States, Tian Lu, Raghu Santanam, Benjamin Shao

Effective collaboration between humans and artificial intelligence (AI) involves leveraging complementary skills and expertise, requiring humans to contribute critical thinking, contextual sensitivity, and judgment in final decisions. This becomes particularly salient when others are impacted by AI generated choice. According to behavioral decision-making literature, affect plays a crucial element in enhancing problem-solving capabilities. Understanding this effect in decision environments assisted by AI warrants further consideration. We introduce empathic algorithm delegation (EAD) to describe the fundamental empathic considerations decision-makers should take during AI decision delegation (interpersonal perspective-taking, contextual understanding, and final decision responsibility). We investigate the implications of this framework on deliberation behaviors and decision accuracy using a Chinese-based micro-lending context. We find the individual components of this framework, separately and together, yield significant improved human-AI synergies in contexts where choices have significant interpersonal impacts.

4 - Generative Policy Learning via Diffusion Models

Zekai Fan, Carnegie Mellon University, Pittsburgh, PA, United States, Michael Lingzhi Li, Shixiang Zhu

Data-driven decision-making tools are widely used in critical applications but often require decision-makers to explicitly define target value functions and constraints. As Michael Polanyi observed, "We can know more than we can tell," highlighting the challenge of formalizing tacit knowledge and preferences. To address this, we propose a framework called "generative policy learning," which leverages a conditional diffusion model to produce diverse, near-optimal, and potentially high-dimensional policies for any chosen value function. Our approach introduces a re-weighting scheme for training diffusion models based on the probability of each sample being near-optimal. This framework allows decision-makers to focus on primary objectives and apply their tacit knowledge to evaluate generated decisions, rather than developing explicit specifications. We demonstrate the effectiveness of this method through extensive synthetic and real-world experiments.

SD70

Regency - 701

Climate Finance and Transition Risk

Invited Session

Energy: Natural Resources

Chair: Felipe Verastegui, Columbia University, 421 W 118th, New York, United States

Co-Chair: Agostino Capponi, Columbia University, New York, NY, United States

1 - Sustainable finance under regulation

Alexandr Kopytov, University of Rochester, Rochester, NY, United States, Shiyang Huang

We build a model analyzing optimal environmental regulation in the presence of socially responsible investors. Investors care about sustainability of their portfolios but cannot fully resolve the pollution externality. Regulations, such as pollution tax and subsidies to clean firms, reduce dirty firms' size but also reshape firms' shareholder compositions. Under the regulations, dirty firms' shareholders become on average less averse to holding polluting shares and hence these firms are less willing to adopt green technologies. We show that pollution can increase with regulation stringency. Optimal regulations do not always fully correct the externality and can deviate from the Pigouvian benchmark.

2 - Climate Regulatory Risks and Corporate Bonds

Lee Seltzer, Federal Reserve Bank of New York, New York, NY, United States, Laura Starks, Qifei Zhu

Investor and policymaker concerns about climate risks suggest these risks should affect the risk assessment and pricing of corporate securities, particularly for firms facing stricter regulatory enforcement. Using corporate bonds, the authors find support for this hypothesis. Employing a shock to expected climate regulations, they show climate regulatory risks causally affect bond credit ratings and spreads. A structural credit model indicates that the increased spreads for high carbon issuers, especially those located in stricter regulatory environments, are driven by changes in firms' asset volatilities rather than asset values, highlighting that regulatory uncertainty affects security pricing. The results have important implications for policy-making.

3 - Green Intermediary Asset Pricing

Maxime Sauzet, Boston University, Boston, MA, United States

Can environmentally-minded investors impact the cost of capital of green firms even when they invest through financial intermediaries? To answer this and related questions, I build an equilibrium intermediary asset pricing model with three investors, two risky assets, and a riskless bond. Specifically, two heterogeneous retail investors invest via a financial intermediary who decides on the portfolio allocation that she offers between a green and a brown equity. All investors can tilt towards the green asset, beyond pure financial considerations. Perhaps surprisingly, the green retail investor can have significant impact on the pricing of green assets, even when she invests via an intermediary who does not tilt: a sizable green premium can emerge on the equity of the green firm. This good news comes with important qualifications, however: the green retail investor has to take large leveraged positions in the portfolio offered by the intermediary, her strategy must be inherently state-dependent, and economic conditions or the specification of preferences can overturn or limit the result. When the financial intermediary decides (or is made) to tilt instead, the impact on the green premium is substantially larger, although it is largest when preference are aligned with retail investors. I also study what happens when the green retail investor does not know the weights in the portfolio offered by the intermediary, the potential impact of greenwashing, and the effect of portfolio constraints. Taken together, these findings highlight the central role that financial intermediaries play in channeling financing (or not) towards the green transition.

4 - Sustainable Investing Strategies with Tradable Assets

Felipe Verastegui, Columbia University, New York, NY, United States, Agostino Capponi

Many sustainable investing strategies employ green-screening, setting pollution thresholds that firms must meet to receive financing from socially responsible investors. However, if the aim is to cut emissions, this method may be ineffective, as firms can change the composition of their asset portfolios without adopting cleaner technologies. To address this, we introduce a competitive equilibrium model featuring heterogeneous firms and a single sustainable fund, which aims to promote decarbonization across a finite asset pool. We contrast green-screening with conditional decarbonization, where the fund mandates a minimum emission reduction. Firms make ex-ante commitments to effort allocations and engage in real asset trades, to then face a random match process which determines whether they secure funding for their operations or not. We show that compliance decisions exhibit strategic complementarity when the size of the sustainable fund is small. When including asset trades, we find that the presence of a sustainable fund can lead to carbon leakage, where assets shift from compliant to non-compliant firms. If transaction costs are small relative to the cost of clean technology, trade decisions over pollutive assets also exhibit strategic complementarity. While green-screening concentrates leakage in carbon-intensive assets, conditional decarbonization mitigates this risk, reverting the relative ordering of compliance costs. We demonstrate that conventional sustainable investment targets fail to minimize total emissions under worst-case scenarios of ex-post firm-investor matches.

SD71

Regency - 702

Generative Artificial Intelligence for Physics Science and Acceleration

Invited Session

Data Mining

Chair: Bo Shen, New Jersey Institute of Technology, Newark, NJ, United States

Co-Chair: Raghav Gnanasambandam, Virginia Tech, Blacksburg, VA, 24060, United States

1 - Data-Analytic Opportunities and Challenges in Solar Eruption Forecasting

Victor Verma, University of Michigan, Ann Arbor, MI, United States, Yang Chen

In recent years, there has been a growing awareness of space weather impacts on critical infrastructure in the civilian, commercial, and military sectors. The interest will continue to grow as we gain a better understanding of the physical processes of the Sun and their effects here on Earth and in space itself. Among all space weather events, the solar flare is a relatively intense, localized emission of electromagnetic radiation in the Sun's atmosphere. Flares occur in active regions and are often, but not always, accompanied by coronal mass ejections, solar particle events, and other eruptive solar phenomena. In recent years, machine learning models have been employed to forecast solar eruptions, including strong solar flare events.

In this talk, I will discuss data-analytic opportunities and challenges in solar eruption forecasting, given the availability of massive solar imaging data and other relevant data products. Scientifically, it is essential to (a) identify the solar active regions that have high potential to erupt in an automated fashion, (b) extract features from observed solar images using principled statistical algorithms, (c), most importantly, provide a probabilistic forecast for the eruption time, magnitude and magnetic field configuration, and (d) facilitate new understandings of the mechanism/physics of disruptive space weather events. In particular, I will present one of our novel statistical models, the Tensor Gaussian Process with Contraction, for solar flare forecasting, combining data from various types and sources. I will conclude with ongoing and future work on operational solar eruption forecasting.

2 - Generation of He I 1083 Nm Images from HA Images by Deep Learning

Marco Marena, New Jersey Institute of Technology, Newark, NJ, United States, Qin Li, Haimin Wang, Bo Shen

This research introduces an innovative method for augmenting the Helium 10830 Å dataset using advanced deep learning techniques, specifically through the Pix2PixHD conditional Generative Adversarial Network (cGAN). The primary objective of this study is to overcome the temporal limitations of the existing Helium 10830 Å dataset, which spans from 2011 to 2020. This augmentation leverages the abundant H-alpha (Hα) spectral line data as a foundational predictor for generating Helium 10830 Å images. The introduction of Helium 10830 Å observations marked a transformative era in solar physics by providing unprecedented detail of the sun's surface and atmospheric behavior. Our work aims to extend the temporal range and enhance the utility of the Helium 10830 Å dataset by integrating it with an extensive repository of Hα data. We employed the Pix2PixHD model, training it on correlated pairs of Hα and Helium 10830 Å images, to facilitate an accurate conversion from Hα data into the Helium 10830 Å domain. This approach not only expands the scope of the Helium 10830 Å dataset but also significantly improves our understanding of solar atmospheric activities over an extended timeline. The generated Helium 10830 Å images demonstrate a strong correlation with actual observational data, with a correlation coefficient (CC) of 0.91 and a Root Mean Square Error (RMSE) of 8.62. These results attest to the model's ability to intricately replicate features of the solar atmosphere. Consequently, this advancement heralds a new epoch in solar physics and data generation methodologies, emphasizing the study's significance within the scientific community.

3 - Automatic Fault Characterization of Multiple Signals Using Bayesian Tensor Modeling with Semiconductor Inspection Data

Qing Zou, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Andi Wang, Jian Li

Drawing inspiration from historical semiconductor inspection datasets, this paper introduces a novel Bayesian modeling framework tailored for unsupervised learning tasks involving high-dimensional, smooth, and low-rank tensor data. To enhance the detection of abnormal events, our proposed model integrates a smooth and sparse components tensor decomposition method, employing Gaussian process priors and sparsity priors, respectively. Additionally, we incorporate Bayesian CP decomposition with automatic rank determination to facilitate efficient modeling of high-dimensional tensor data. Within this innovative Bayesian framework, we present an efficient Gibbs sampling approach to tackle the model inference problem. Both numerical experiments and a real-world case study utilizing semiconductor datasets demonstrate the efficacy of our proposed methodology.

4 - Direct Ink Writing Process Analysis from a Multi-Stage System Perspective

Wuyang Chen, University of Georgia, Athens, GA, United States, Chenyu Xu, Chi Zhou, Hongyue Sun

This research focuses on the multi-stage analysis of Direct Ink Writing (DIW) 3D printing. DIW can be used for the printing of various functional materials, which often exhibit complex properties that affect the subsequent printing and drying processes. Traditional physical modeling usually cannot handle such dynamic interactions. To address this challenge, we adopt a data-driven approach, collecting extensive data across all stages, from the initial material properties to real-time sensing and imaging during printing, and detailed evaluations of the final products like 3D scans and thermal properties. We analyze and decode the interdependencies among the DIW stages, which facilitates the comprehensive understanding and optimization of the DIW process.

SD72

Regency - 703

Sequential Learning: Optimization and Prediction Under Uncertainty

Invited Session

Data Mining

Chair: Imtiaz Ahmed, West Virginia University, Morgantown, WV, United States

Co-Chair: Hamed Khosravi, West Virginia University, Morgantown, WV, United States

1 - Active Learning of Piecewise Gaussian Process Surrogates

Chiwoo Park, University of Washington, Seattle, WA, United States, Robert Waelder, Bonggwon Kang, Benji Maruyama, Soondo Hong, Robert Gramacy

Active learning of Gaussian process (GP) surrogates has been useful for optimizing experimental designs for physical/computer simulation experiments, and for steering data acquisition schemes in machine learning. Here, we present a method for active learning of piecewise, Jump GP surrogates. Jump GPs are continuous within, but discontinuous across, regions of a design space, as required for applications spanning autonomous materials design, configuration of smart factory systems, and many others. Although our active learning heuristics are appropriated from strategies originally designed for ordinary GPs, we demonstrate that additionally accounting for model bias, as opposed to the usual model uncertainty, is essential in the Jump GP context. Toward that end, we present an estimator for bias and variance of Jump GP models. Illustrations, and evidence of the advantage of our proposed methods, are provided on a suite of synthetic benchmarks, and real-simulation experiments of varying complexity.

2 - Lnucb-Ta: Linear-Nonlinear Hybrid Bandit Learning with Temporal Attention

Imtiaz Ahmed, West Virginia University, Morgantown, WV, United States, Hamed Khosravi

In this study, we present LNUCB-TA, a novel hybrid contextual multi-armed bandit algorithm that effectively combines linear and non-linear estimation strategies to enhance decision-making accuracy in stochastic contextual bandit problems. This model introduces significant advancements over the vanilla Linear Upper Confidence Bound (LinUCB) algorithm by integrating an adaptive k-nearest neighbors (k-NN) method alongside an attention-based exploration mechanism, designed to enhance the precision of reward estimations and adaptability in complex environments.These novel features adjust the confidence bands and reward estimations for each option or "arm," by utilizing learned data-driven similarities among the feature vector and reward history associated with each arm. This significantly reduces the likelihood of selecting non-optimal arms. By leveraging both temporal and spatial complexities, our model transitions from static to dynamic exploration rates, effectively balancing exploration and exploitation in real time.We demonstrate the model’s efficiency and robustness through a theoretical analysis, proving that the regret bound is sub-linear. Furthermore, empirical results demonstrate that LNUCB-TA outperforms several state-of-the-art bandit models.

3 - Optimal Resource Allocation in a Collaborative Production System at Subtask Level

Congfang Huang, UW Madison, Madison, WI, United States

This talk will introduce a resource allocation problem in a serial manufacturing production line consists of multiple workstations. The available resources include human workers, independent robot and collaborative robots. At the subtask level, we consider the precedence relationship between different subtasks. The objective is to improve the throughput, reduce the economic cost and release ergonomic strain. A reinforcement learning guided evolutionary algorithm is proposed to solve the problem and a real world case study will be illustrated.

4 - Attention Enhanced Damage Prediction in Electrode Materials

Quan Zeng, Georgia Institute of Technology, Atlanta, GA, United States

In this talk, we report our recent progress in damage prediction based on application of a recurrent neural network model to fracture videos. Specifically, we are able to apply physically meaningful stress gradients as attention guidance and to accurately predict damage propagation for the next image frame. In this research, the damage and stress videos were simulated using phase-field model, but should the X-ray microscopy data be available, we are confident that the resulting method is applicable as well.

SD73

Regency - 704

Emerging Trends in Interpretable Machine Learning

Invited Session

Data Mining

Chair: Lesia Semenova, Microsoft Research, NY, 27713, United States

1 - Recent Work in Dimension Reduction for Data Visualization

Cynthia Rudin, Duke University, Durham, NC, United States

Dimension reduction (DR) for data visualization is unsurpassed in its ability to provide insight from data. I will discuss recent work from the Interpretable Machine Learning Lab in dimension reduction and applications.

2 - Interpreting CLIP with Sparse Linear Concept Embeddings (SpLiCE)

Usha Bhalla, Harvard University, Cambridge, MA, United States, Alex Oesterling, Suraj Srinivas, Flavio Calmon, Himabindu Lakkaraju

CLIP embeddings have demonstrated remarkable performance across a wide range of computer vision tasks. However, these high-dimensional, dense vector representations are not easily interpretable, restricting their usefulness in downstream applications that require transparency. In this work, we empirically show that CLIP's latent space is highly structured, and consequently that CLIP representations can be decomposed into their underlying semantic components. We leverage this understanding to propose a novel method, Sparse Linear Concept Embeddings (SpLiCE), for transforming CLIP representations into sparse linear combinations of human-interpretable concepts. Distinct from previous work, SpLiCE does not require pre-defined concept labels and can be applied post hoc to unlabelled data. Through extensive experimentation with multiple real-world datasets, we validate that the representations output by SpLiCE can explain and even replace traditional dense CLIP representations, maintaining equivalent downstream performance while significantly improving their interpretability. We also demonstrate several use cases of SpLiCE representations including detecting spurious correlations, model editing, and quantifying semantic shifts in datasets.

3 - The Dataset Multiplicity Problem: How Unreliable Data Impacts Predictions

Anna Meyer, University of Wisconsin - Madison, Madison, WI, United States, Aws Albarghouthi, Loris D'Antoni

We introduce dataset multiplicity, a way to study how inaccuracies, uncertainty, and social bias in training datasets impact test-time predictions. The dataset multiplicity framework asks a counterfactual question of what the set of resultant models (and associated test-time predictions) would be if we could somehow access all hypothetical, unbiased versions of the dataset. We discuss how to use this framework to encapsulate various sources of uncertainty in datasets' factualness, including systemic social bias, data collection practices, and noisy labels or features. We show how to exactly analyze the impacts of dataset multiplicity for a specific model architecture and type of uncertainty: linear models with label errors. Our empirical analysis shows that real-world datasets, under reasonable assumptions, contain many test samples whose predictions are affected by dataset multiplicity. Furthermore, the choice of domain-specific dataset multiplicity definition determines what samples are affected, and whether different demographic groups are disparately impacted. Finally, we discuss implications of dataset multiplicity for machine learning practice and research, including considerations for when model outcomes should not be trusted.

4 - Revolutionizing Clinician Diagnosis of Seizure and Seizure-like EEG Patterns using Interpretable AI

Alina Barnett, Duke University, Durham, NC, United States

We designed a first-of-its-kind AI assistive tool to prevent serious brain injury and death for critical care patients by illuminating important patterns in their electroencephalogram (EEG) signals. Our solution addresses the societal issue of low clinician trust in AI models by inventing novel technical solutions to elucidate the decision-making of famously opaque neural networks. With our model, clinician classification accuracy improves substantially. This technology can greatly improve patient care both in well-served areas and especially in poorly-served areas.

5 - Provably-Optimal Sparse Trees

Rui Zhang, Duke University, Durham, NC, United States, Rui Xin, Cynthia Rudin

Decision trees are one of the oldest form of AI models. Within its vast literature, a large number of work rely on heuristic (or greedy) algorithms to produce tree models, which risk yielding sub-optimal models. Researchers have made recent success on fully optimized classification trees, but they fail to extend it to more generalized tasks (regression, survival analysis and multi-label learning). Our work fill this gap by providing provably-optimal sparse trees for more generalized tasks in a fast and efficient way.

SD74

Regency - 705

Long-Term Power System Expansion Planning with New Renewable Fuels

Invited Session

ENRE: Electricity

Chair: Enzo Sauma, Pontificia Universidad Catolica de Chile, Santiago, Chile

1 - Impact of on-grid ammonia production on power system expansion planning under European Union definitions of renewable fuels

Cristóbal Zuazagoitia, Pontificia Universidad Católica de Chile, Santiago, Chile, Enzo Sauma

The Commission Delegated Regulation (EU) 2023/1184, supplementing Directive (EU) 2018/2001 of the European Parliament and of the Council was published in the Official Journal of the European Union in June 2023. This regulation establishes a methodology, among other statements, describing specific rules for considering electricity taken from the grid as fully renewable to produce renewable liquid and gaseous transport fuels of non-biological origin (RFNBO). In this work, we investigated the impact of on-grid ammonia production on the long-term expansion planning of the Chilean power system, aligning with European Union definitions of RFNBO for future exports. We propose a comprehensive model formulated as a Mixed-Integer Linear Program (MILP) to minimize investment and operational costs, while considering technical constraints such as Kirchhoff's Voltage Law (KVL). The incorporation of renewable energy quotas and/or maximum CO2 emission factor per period results in a generation mix and installation of new transmission lines that significantly vary temporally and spatially from the base case. By modeling ammonia production as a partially flexible load, we explore its impact on the entire power system, including systemic costs, line congestion, curtailments, and seasonal production profiles.

2 - Long-term expansion planning of power systems with the integration of green ammonia under different economic and regulatory scenarios for the year 2060: the case of Chile

Francisca Meriño Fariña, Pontificia Universidad Catolica de Chile, Santiago, Chile, Enzo Sauma

Climate change is a significant global challenge. Due to the Chile's large renewable-energy potential, green hydrogen emerges as an energy vector that can support the decarbonization of the various energy-demanding sectors. Since the transport of hydrogen is complex and expensive, ammonia emerges as one of the best options for its transportation. This work proposes a long-term expansion planning model of the Chilean National Electric System, which integrates hydrogen and green ammonia to the year 2060, addressing the urgent need for decarbonization. This long-term expansion planning model of the power system for the year 2060 optimizes the installation of ammonia production plants subject to some technical constraints.

Results suggest that the high penetration of renewable energy and the use of storage systems will make viable an economic production of hydrogen and green ammonia by the year 2060. We also analyze the main interactions between the optimal electricity system plan and the production of these clean fuels under various economic and regulatory scenarios

3 - Evaluation of the economic and environmental impacts of implementing different public policy mechanisms to accelerate the development of the H_2 industry in Chile

Enzo Sauma, Pontificia Universidad Catolica de Chile, Santiago, Chile, Andrés Pereira, Francisco Manríquez, Sonia Vera

We first formulate and implement a model that co-optimizes, from a system-wide perspective, the expansion of the H_2 industry and the expansion of the associate power system, when considering the main technical features of both the power system operation and the H_2 production. This model is implemented in the "Platform for Long-Term Planning Analysis of National Electric System" (PLANES) developed. Using PLANES, we evaluate the economic and environmental impacts of implementing different public policy mechanisms aimed at encouraging and accelerating the development of the H_2 industry in Chile. Our results indicate that aligning electrolyzer operations with the hourly availability of renewable resources (i.e., encouraging a coordination between the hydrogen system operation and the electricity system operation) can lead to a reduction in the operation hours of fossil fuel power plants, decreased reliance on Battery Energy Storage Systems (BESS), and curtailed electricity, resulting in a lower global total cost of electricity.

4 - What duality theory tells us about giving the market operator the authority to dispatch energy storage

Yuzhou Jiang, Verse Inc., San Francisco, CA, United States, Ramteen Sioshansi

Energy storage has becoming an important participant in modern electricity markets. The question of which entity should have the authority to make operational decisions for energy storage has been ongoing. Some stakeholders raise concerns that market operators' independence can be undermined if they make operational decisions for energy storage. The rationale that underlies this concern is that operating energy storage can affect the balance of the system and price formation. We formulate optimal-power-flow problem as well as its dual problem, and demonstrate that having market operators make operational decisions for energy storage does not change its fundamental nature. We show that if market operators co-optimize the operation of energy storage with that of generators and transmission, the optimal-power-flow problem yields short-run and long-run efficiency. These findings are analogous to those for having market operators co-optimize transmission use with generator dispatch. Our work suggests that concerns around giving market operators the authority to dispatch energy storage are misplaced.

SD75

Regency - 706

Climate Uncertainty and Healthcare

Invited Session

ENRE: Energy-Climate

Chair: Chunjie Zhao, Clark University, LA PUENTE, United States

Co-Chair: Tianyi Yang, City of Hope, Monrovia, CA, United States

1 - From Policy to Power Savings: Role of Smart Technology Installations in Reducing Household Electricity Bills

Chunjie Zhao, Clark University, Worcester, MA, United States

Smart technologies play one important role in improving quality of life and fostering sustainable development. This study investigates the factors influencing smart thermostat adoption, with a specific focus on the different policy measures that encourage such adoption. Moreover, it explores how these policies potentially narrow the adoption gap between urban and rural regions. Through the logistic regression, we identified a positive relationship between the implementation of policies and the promotion of smart thermostat adoption, with the tax rebate policy being the most effective in mitigating urban-rural adoption disparities. These findings contribute to the broader discourse on sustainable development, technology diffusion, and equitable socio-economic progress.

2 - Mitigating the Healthcare Impacts of Climate Change through Sustainable Energy and Environmental Policies

Grace Babalola, Binghamton University (SUNY), Johnson City, NY, United States, Timilehin Halim, Lyndia Yusuf, Oluwatunmise Halim, Simeon Ogunbunmi

Climate change poses significant risks to public health, with potential consequences such as increased air pollution, water scarcity, vector-borne diseases, and extreme weather events. This work explores how sustainable energy and environmental conservation policies can help mitigate these healthcare impacts by reducing greenhouse gas emissions, promoting clean energy sources, and fostering a more resilient environment – delving into the link between climate change and healthcare challenges, including respiratory diseases, heat-related and the spread of infectious diseases.

The role of sustainable energy in mitigating these impacts is highlighted, focusing on transitioning to renewable sources like solar and wind power, which improves air quality and reduces air pollution-related respiratory issues. Energy efficiency measures in healthcare are also discussed, as they contribute to lowering the carbon footprint and promoting sustainability.

Environmental conservation strategies, such as reforestation, sustainable agriculture, and water conservation, are examined for their potential to mitigate the effects of climate change and protect public health. Additionally, it explores the benefits of circular economy practices and waste management strategies in reducing environmental pollution and associated healthcare costs.

Furthermore, the importance of interdisciplinary collaborations between healthcare professionals, policymakers, energy experts, and environmental advocates is emphasized, underscoring the need for comprehensive policy frameworks and incentives to drive sustainable energy and environmental conservation efforts illustrated with real-world success stories.

Ultimately, this work provides a forward-looking perspective on the potential of sustainable energy and environmental policies to address climate uncertainty and its healthcare implications, offering recommendations for stakeholders to promote and support these vital initiatives.

3 - Adaptive Hidden Factor Structured Bayesian Networks and the Applications

Peng Jiang, Sichuan University, Chengdu, China, People's Republic of

Bayesian networks are appropriate candidates for systematic modeling with complex causalities and decision-making under uncertainties. However, multi-dimensional influencing variables incorporated in the networks would result in non-robust parameter estimations and expensive computations. To simplify the number of parameters and enhance parameter estimations, this study proposes a novel Adaptive Hidden Factor Structured Bayesian Network (AHFSBN) whose parameters and hidden structure are adaptively learned through multivariate spatial-temporal data in a semi-supervised learning manner. Theoretical analyses are given regarding the network structure, the parameter space, and the convergence of the related parameter learning algorithm. Then, the AHFSBN has been applied to health risk prevention modeling and risk management of emerging contaminants including cyanotoxins and antimicrobial resistance genes in natural aquatic environments. These real-world cases have demonstrated that the AHFSBN can provide intelligent data-driven decision support and offer several managerial insights.

4 - Echo State Networks For Short-Term Regulation Reserve Deployment Forecasting

Hugo Santarem de Araujo, Cornell University, Ithaca, NY, United States, Jose Daniel Lara, Jacob Mays, Bri-Mathias Hodge

Hybrid merchants in a competitive wholesale electricity market may seek to minimize their operational costs by storing low marginal-cost resources to provide energy or ancillary services. To provide frequency-regulation services, merchants employ past data and all sorts of models and heuristics to decide how to position their storage resources to fulfill their binding offer. Despite the growing proliferation of storage, the highly irregular nature of the frequency-regulation services data makes the prediction task rather challenging and the operation of batteries risky. In response to this challenge, we employed Echo State Networks to carry out the predictions of frequency-regulation services aiming at integrating the forecasts to the operation of the day-ahead market in the Electricity Reliability Council of Texas and the California ISO. Adopting a recursive framework, predictions refreshed every 5 minutes for about 22 days employing classical and ML methods. Results show that the proposed approach successfully captures the signal in the data as error metrics can be up to 10 times more competitive than the next best option.

5 - Compilation, Interpretation, and Integration of Wasted Food within Optimization-Based Models

Sauleh Siddiqui, American University, Washington, DC, United States

We present an expansion of US FEAST, a partial-equilibrium model of the US Food System, to include Wasted Food dynamics for projecting future outcomes. The food system represents an integrated series of systems that constitutes various stages of production, processing, distribution, and consumption. Each segment in the supply chain is tightly interconnected where alterations in one area reverberate throughout the entire system. Moreover, it intersects with other critical infrastructures like energy and transportation and impacts economics, public health, and the environment. Optimization models are frequently utilized in the mathematical representation of food system models and have played a crucial role in analyzing various dimensions of the food system. We explore methods for integrating the latest data and dynamics related to wasted food within food system models. Such types of models play a crucial role in quantifying the effects of wasted food on various parts of the supply chain. The model presented here offers insights into the potential impacts of wasted food reduction on food production, flow, and consumption.

SD76

Regency - 707

Energy Systems Modeling and Optimization II: EV Integration Under Uncertainty

Invited Session

ENRE: Other Energy

Chair: Tanveer Bhuiyan, The University of Texas at San Antonio, San Antonio, TX, 78249, United States

1 - Adoption of EVs in Municipalities in Ma

Edgar Campos, Bentley University, Waltham, MA, United States, Zana Cranmer

We conducted a study in Massachusetts to understand why some towns have lower rates of electric vehicle (EV) adoption than others. Our goal was to identify the factors that influence EV adoption, including various demographic factors, EV charging and public transit infrastructure, and environmental attitudes. To achieve this, we used a multivariate linear model and found that the only factors that were not significant were our infrastructure variables. We found that income, voting results, urban/rural measurements of the town, homeownership, and environmental attitudes were significant in explaining EV adoption rates. Our findings have important implications for policymakers who are interested in promoting EV adoption, as outlined in the Massachusetts Clean Energy and Climate Plan.

2 - A Stochastic Game-Theoretic Optimization Approach for Managing Local Electricity Markets with Electric Vehicles and Renewable Sources

Sayed Hamid Hosseini Dolatabadi, The University of Texas at San Antonio, San Antonio, TX, United States, Tanveer Bhuiyan, Yang Chen

Local electricity markets (LEMs) are pivotal in optimizing active distribution systems, driven by renewable energy sources (RESs), and electric vehicles (EVs). Our study presents a stochastic bi-level model for managing LEMs, where the local transaction center (LTC) and other market participants (agents) (load aggregator (LA), charging station (CS), and their lower-level prosumers) aim to maximize their profits in a competitive game-theoretic structure. To ensure grid stability and profitability, LTC as the leader seeks to design a robust dynamic pricing strategy under uncertainty, whereas LA and CS, as followers, decide on the energy transaction amounts under uncertainty in their generation and consumption. We address several novel aspects of LEMs, such as heterogeneous behaviors of entities, uncertainties in renewable generation and load profiles, and the consumption profile of the EV fleet. We offer two solution methodologies including an exact centralized and a hybrid decentralized solution approach. Based on a realistic case study, we provide key insights into the effect of decision-making on the market dynamics, the benefit of our proposed dynamic pricing, the benefit of modeling stochasticity in generation and load profiles, and the performance of different solution approaches. Specifically, numerical results demonstrate that increasing the capacity of LTC's energy storage and its initial state of charge increases LTC's profit by 255%. Results also demonstrate that modeling the LEM under uncertainty as a two-stage stochastic program provides a robust pricing decision resulting in a 433% higher profit compared to the deterministic modeling that ignores the uncertainties in the key parameters.

3 - Optimizing E-commerce Electric Delivery Vehicle Charger Locations under Stochastic Demand

Taner Cokyasar, Argonne National Laboratory, Lemont, IL, United States, Omer Verbas, Tanveer Bhuiyan

As the demand for online shopping surges, the need for efficient last-mile delivery solutions becomes increasingly important, with electric vehicles (EVs) emerging as a sustainable option. However, the stochastic nature of delivery demand and the resulting inconsistent routing scheme across operational days poses a significant challenge in determining optimal charger locations to ensure timely and uninterrupted service. This study proposes a novel methodology that integrates stochastic location modeling and queueing theory to pinpoint optimal charger locations. Using a mixed-integer non-linear program, the problem is formulated, and an exact solution method employing cutting planes as lazy constraints is devised. The model aims to minimize costs associated with stations, chargers, detours, and queueing and charging times, thereby determining the optimal location, type, and quantity of EV chargers. The exact solution method contributes to the existing literature on stochastic location modeling. To enhance scalability, the study develops clustering heuristics and simulated annealing metaheuristic approaches. The research outcomes contribute to the advancement of understanding in sustainable urban logistics systems, providing actionable insights for businesses and policymakers aiming to improve the efficiency and sustainability of e-commerce delivery operations amid shifting consumer behaviors and environmental considerations.

4 - Planning of Networked Microgrids for Hydrogen-powered Vehicles: A Study on Induced Refueling Demand

Bo Zeng, University of Pittsburgh, Pittsburgh, PA, United States, Xunhang Sun, Xiaoyu Cao, Qiaozhu Zhai, Tamer basar, Xiaohong Guan

A planning model is presented for networked microgrids for hydrogen-powered vehicles, which considers the impact of induced demand.

5 - Shared Electric Vehicle Fleet for Stable and Fair Ride-Sharing Service

YANTAO HUANG, Argonne National Laboratory, Lemont, IL, United States, Hui Shen, Krishna Murthy Gurumurthy, Joshua Auld

Centralized ride-sharing service provides optimized fleet performance by picking the best candidate riders for vehicles, but with no considerations to rider preferences. This study explores how energy consumption and network performance will change when a ride-sharing service is provided by an electric vehicle (EV) fleet, which follows a vehicle and passenger matching rule that respects the preference considerations from both riders and fleet. That means fair considerations of the routing and charging on the fleet side, and wait times, seating availability and driver ratings on the riders' side, which eventually matches vehicles and riders to achieve a preference equilibrium. This paper advances the traditional stable and fair matching algorithm (SFMA) from one-to-one matching to many-to-one matching, extending the use case from transportation network companies' (TNC's) ride-sourcing service to ride-sharing service. This new many-to-one SFMA (MSFMA) will be evaluated using large-scale agent-based simulator POLARIS (Auld et al., 2016), which is capable of simulating millions of trips over a 24-hour simulation period in large metro areas. An optimization-based matching algorithm from Alonso-Mora (2017), often used as a benchmark, will give a baseline of vehicle-related events (e.g., routing, charging, and wait times), riders' experience (e.g., wait time) and network performance (e.g., average vehicle occupancy and road congestion). Comparisons would detail changes between an optimization-driven matching method and an equitable matching method for the electric ride-sharing service, in order reveal how two-sided market situation differs from centralized optimal fleet operations in terms of energy consumption and network performance.

SD77

Regency - 708

Computation in Conic Programming

Invited Session

Computing Society

Chair: Boshi Yang, Clemson University, Clemson, SC, United States

Co-Chair: Hao Hu, Clemson University, Clemson, SC, United States

1 - The Maximum Singularity Degree for Linear and Semidefinite Programming

Hao Hu, Clemson University, Clemson, SC, United States

The singularity degree plays a crucial role in understanding linear and semidefinite programming, providing a theoretical framework for analyzing these problems. It is defined as the minimum number of facial reduction (FR) steps needed to reach strict feasibility for a convex set. On the other hand, the maximum singularity degree (MSD) is the maximum number of steps required. Recent progress in the applications of MSD has motivated us to explore its fundamental properties. For semidefinite programming, we establish a necessary condition for an FR sequence to be the longest. Additionally, we propose an upper bound for MSD, which can be computed more easily. By leveraging these findings, we prove that computing MSD is NP-hard. This complexity result complements the existing algorithms for computing the singularity degree found in the literature. For linear programming, we provide a characterization for the longest FR sequences, which also serves as a polynomial-time algorithm for constructing such a sequence. In addition, we introduce two operations that ensure the longest FR sequences remain the longest. Lastly, we prove that MSD is equivalent to a novel parameter called the implicit problem singularity.

2 - Revisiting Spectral Bundle Methods: Primal-Dual (Sub)Linear Convergence Rates

Lijun Ding, Texas A&M University, College Station, TX, United States, Benjamin Grimmer

The spectral bundle method proposed by Helmberg and Rendl is well established for solving large-scale semidefinite programs (SDP) thanks to its low per iteration computational complexity and strong practical performance. In this paper, we revisit this classic method showing it achieves sublinear convergence rates in terms of both primal and dual SDPs under merely strong duality, complementing previous guarantees on primal-dual convergence. Moreover, we show the method speeds up to linear convergence if (1) structurally, the SDP admits strict complementarity, and (2) algorithmically, the bundle method captures the rank of the optimal solutions. Such complementary and low rank structure is prevalent in many modern and classical applications. The linear convergent result is established via an eigenvalue approximation lemma which might be of independent interest. Numerically, we confirm our theoretical findings that the spectral bundle method, for modern and classical applications, speeds up under these conditions. Finally, we show that the spectral bundle method combined with a recent matrix sketching technique is able to solve an SDP with billions of decision variables in a matter of minutes.

3 - Cone product reformulation for global optimization

Theodoros Koukouvinos, MIT, Cambridge, MA, United States, Dimitris Bertsimas, Dick den Hertog, Danique de Moor, Jianzhe Zhen

In this paper, we study nonconvex optimization problems involving sum of linear times convex (SLC) functions as well as conic constraints belonging to one of the five basic cones, that is, linear cone, second order cone, power cone, exponential cone, and semidefinite cone. By using the Reformulation Perspectification Technique, we can obtain a convex relaxation by forming the perspective of each convex function and linearizing all product terms with newly introduced variables. To further tighten the approximation, we can pairwise multiply \rev{(parts of)} the conic constraints. In this paper, we analyze all possibilities of multiplying conic constraints. Especially the results for the cases in which a power cone or an exponential cone is involved are new. Particularly noteworthy are the novel results involving the power cone and exponential cone. We delineate methods for deriving new, valid linear and second-order cone inequalities for pairwise constraint multiplications involving these cones, thereby enhancing the strength of the approximation. Numerical experiments on a quadratic optimization problem over exponential cone constraints and on a robust palatable diet problem over power cone constraints, demonstrate that including additional inequalities generated from the proposed pairwise multiplications improve the approximation. Moreover, when incorporated in branch and bound the global optimal solution of the original nonconvex optimization problem can often be obtained faster than BARON.

SD78

Regency - 709

Recent Advances in Decision Rules for Stochastic and Robust Optimization

Invited Session

Computing Society

Chair: Zolykha Rezaei, Texas Tech University, Lubbock, TX, United States

1 - A Group Sparsity-Based Stochastic Gradient Descent Algorithm for Beam Angle Optimization in Impt

Hadis Moazami Goudarzi, University of Houston, Houston, TX, United States, Gino Lim, Radhe Mohan, Wenhua Cao

This study introduces a new beam angle optimization (BAO) approach for Intensity-Modulated Proton Therapy (IMPT) by considering biological aspects of cancer treatment. Unlike conventional BAO methods that focus mainly on physical dose, our approach integrates proton therapy's biological effects incorporating Linear Energy Transfer (LET) into BAO to improve treatment effectiveness. This integration is critical as LET significantly affects the biological effectiveness of the treatment, promoting more personalized cancer therapies. We will present (1) a new group sparsity-based model for BAO, (2) a mini batch semi stochastic gradient descent algorithm (mS2GD PGCD) to enhance the speed of the BAO process without sacrificing dosimetric accuracy, and (3) a comparative analysis showing that incorporating LET into BAO with the LETinc mS2GD PGCD algorithm yields better OAR sparing than the LETinc FISTA algorithm and clinical angles. Our findings reveal that LETinc mS2GD PGCD not only enhances the overall treatment efficacy but also improves the sparing of organs-at-risk (OAR) compared to LETinc FISTA, emphasizing the importance of computational efficiency and biological impact in treatment plan optimization. Specifically, LETinc mS2GD PGCD achieves better control of mean and maximum doses across critical structures such as the brainstem and spinal cord, with reductions in cLET×D that suggest a lower risk of radiation-induced complications. These results underscore the superiority of the LETinc mS2GD PGCD algorithm in balancing the need for effective tumor control while minimizing damage to surrounding healthy tissues.

2 - A parametric column-and-constraint generation algorithm for a class of robust optimization problems under endogenous uncertainty

Jnana Sai Jagana, University of Minnesota, Minneapolis, MN, United States, Qi Zhang

Traditional robust optimization (RO) problems consider exogenous (or decision-independent) uncertainty sets, which can be limiting when the decisions affect the underlying probability distribution or the time of realization of the uncertain parameters. Recently, such endogenous (or decision-dependent) uncertainty has been considered in RO, where the resulting optimization problems have mostly been solved using reformulation techniques [1].

Previous works exploring solution strategies have shown that cutting plane methods also perform competitively for RO problems under exogenous uncertainty [2]. To our knowledge, only one existing work has developed a cutting plane method for RO with decision-dependent uncertainty [3]. The proposed cutting plane method scales better compared to the reformulation approach but is limited to problems where the decisions affecting the uncertainty are discrete.

In this work, we develop a cutting plane algorithm using a column-and-constraint generation (C&CG) strategy for RO under certain polyhedral endogenous uncertainty sets, where the decisions affecting the uncertainty can be both continuous and discrete. The proposed algorithm is applied to several numerical examples, and its performance and scalability compared to the reformulation technique are investigated.

References

1. Nohadani, O., & Sharma, K. (2018). Optimization under decision-dependent uncertainty. *SIAM Journal on Optimization*, *28*(2), 1773-1795.

2. Bertsimas, D., Dunning, I., & Lubin, M. (2016). Reformulation versus cutting-planes for robust optimization: A computational study. *Computational Management Science*, *13*, 195-217.

3. Lappas, N. H. (2018). Robust Optimization for Scheduling Operations under Uncertainty [Doctoral dissertation, Carnegie Mellon University]. ProQuest Dissertations & Theses Global.

3 - Graph-based Learning to Accelerate Two-Stage Robust Optimization in Power Grids

Rohit Dube, Texas A&M University, College Station, TX, United States, Harsha Nagarajan, Natarajan Gautam, Amarnath Banerjee

This talk presents a novel approach to accelerate mixed-integer two-stage adaptive robust optimization (ARO) problems within power grids with load and generation uncertainties by leveraging graph-based deep learning models. The problem involves mixed-integer “here-and-now” decisions for generator commitments and/or network topology variables, as well as an adversarial “wait-and-see” problem with non-convex recourse due to AC power-flow constraints. Traditional ARO methods face challenges in scalability and computational efficiency, particularly in large-scale power grid systems. To address these challenges, we propose a graph-based deep-learning ARO framework. In the framework, the first stage involves selecting decisions for the master problem by solving a mixed-integer program. Subsequently, in the second stage, the worst-case uncertainties corresponding to the first-stage decisions are determined based on the value function of a graph-based learned neural network model. Graph-based methods are used to capture local and global structural sparsity properties of power grids by updating node and edge representations based on relational dependencies within the grid. We demonstrate the effectiveness of our approach against the traditional column-and-constraint generation algorithm, through extensive experiments on realistic power grid datasets. The results showcase significant acceleration with a speed-up of about 10 to 20 times in solving ARO problems compared to traditional methods, with negligible compromise in solution robustness.

4 - K -Adaptability via Orthogonal Partition and Lifting

Zolykha Rezaei, Texas Tech University, Lubbock, TX, United States, Ningji Wei, Eojin Han

In numerous practical scenarios, professionals prefer policies that are not only interpretable but also straightforward to implement, particularly in settings involving sequential decision-making. In this research, we explore the concept of finite adaptability to construct robust optimization policies. More specifically, we address a general setting where uncertainty affects both the objective function and the constraints within the model. Our aim is to construct policies by partitioning the uncertainty realization and assigning a contingent decision to each segment. To effectively solve the robust optimization model under these conditions, we employ two distinct computational methods: Benders' decomposition and the dualization technique for linear functions.

SD99

Flex C

QSR Student Poster Competition

Award Session

Quality, Statistics and Reliability

Chair: Chao Wang, University of Iowa, Iowa City, IA, 52242, United States

Co-Chair: Minhee Kim, University of Florida, Gainesville, FL, 32611, United States

1 - A New View of Neural Network-based Health Index: Connecting Visualization, Prediction, and Operational Decision Making

Zihan Li, University of Florida, gainesville, FL, United States, Minhee Kim

Multisensor signals are now easily accessible for remaining useful life (RUL) prediction and health state monitoring due to the quick advancement of sensor technologies. In order to optimize these signals for improved health condition assessment and RUL prediction, a variety of data fusion approaches are frequently used to generate health indices. However, most of the existing methods hardly consider the interactions between Prognostics and condition-based maintenance (CBM), which may not be effective for practitioners' asset management. To address the issue, we provide a new view of the Health Index that provides Connection Visualization between Prognostics and CBM, Prediction, and Operational Decision-Making Suggestions. In particular, a neural networks-based structure is used, and a novel loss function is developed by taking into account the variability of the generated health index at the failure time as well as its monotonicity. Once the prognostic model is trained, then we search for the best model weights that minimize the maintenance costs along “low loss curves” connecting various trained prognostic model weights. A case study on the degradation of aircraft gas turbine engines is also constructed which shows an improvement in maintenance costs of the proposed method.

2 - Generation of He I 10830 nm Images from H-alpha Images by Deep Learning

Marco Marena, New Jersey Institute of Technology, Newark, NJ, United States

tbd

3 - On Model Compression for Neural Networks: Framework, Algorithm, and Convergence Guarantee

Chenyang Li, New Jersey Institute of Technology, Avenel, NJ, United States

tbd

4 - Strata Design for Variance Reduction in Stochastic Simulation

Jaeshin Park, University of Michigan, Ann Arbor, Ann arbor, MI, United States

tbd

5 - Semi-supervised PARAFAC2 Decomposition for Medical Decision Making

Elif Konyar, University of Florida, Gainesville, FL, United States

tbd

6 - Spatio-Temporal Compartmental Modeling for Customer-Level Power Outages

Shuyi Chen, Carnegie Mellon University, Pittsburgh, PA, United States

tbd

7 - LNUCB-TA: Linear-nonlinear Hybrid Bandit Learning with Temporal Attention

Hamed Khosravi, West Virginia University, Morgantown, WV, United States

tbd

8 - An active learning based Gaussian process surrogate model for low dimensional spectral space projections: applications for in-situ nanosurface imaging via weak acoustic emission sensor signals

Xinchen Wang, Binghamton University, Vestal, NY, United States

tbd

9 - DEN-HMM: Deep Emission Network Based Hidden Markov Model with Time-Varying Observations

Vipul Bansal, University of Wisconsin Madison, Madison, WI, United States

tbd

10 - cancel

Xinchen Wang, Binghamton University, Vestal, NY, United States

tbd

11 - Advanced active learning based smart sensor approach for nanofabrication in-situ characterization

Xinchen Wang, Binghamton University, Vestal, NY, United States

tbd

12 - Computationally Efficient Learning of Artificial Intelligence System Reliability Considering Error Propagation

Fenglian Pan, University of Arizona, Tucson, AZ, United States

tbd

13 - Predicting Uncertainty in Renewable Energy Growth: A Spatio-Temporal Conformal Framework

Wenbin Zhou, Carnegie Mellon University, Pittsburgh, PA, United States

tbd

14 - A Bayesian spike-and-slab sensor selection approach for high-dimensional prognostics

Ye Kwon Huh, University of Wisconsin-Madison, Madison, WI, United States

tbd

15 - Asymptotic Behavior of Adversarial Training Estimator under ℓ_infty-Perturbation

Yiling Xie, Georgia Tech, Atlanta, GA, United States

tbd

16 - A Bayesian Jump Model-based Pathwise Sampling Approach for Online Anomaly Detection

Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States

tbd

17 - Physics-Informed Neural ODE with Heterogeneous Control Inputs (PINOHI) for Quality Prediction of Composite Adhesive Joints

Yifeng Wang, Georgia Institute of Technology, Atlanta, GA, United States

tbd

18 - Resilient Scheduling in Critical Infrastructure Network - A Multi Agent Reinforcement Learning Approach

Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA, United States

tbd

19 - Heterogeneity-Adapted Federated Learning for Collaborative Predictive Analytics in Additive Manufacturing

Anyi Li, Auburn University, Auburn, NY, United States

tbd

20 - Dynamic sensor selections from in-filed units for remote prognostics in the cloud-based system center

Ying Fu, University of Wisconsin-Madison, Madison, WI, United States

tbd

21 - SparseST: Exploiting Data Sparsity in Spatiotemporal Modeling and Prediction

Junfeng Wu, Rensselaer Polytechnic Institute, Troy, NY, United States

tbd

22 - Latent Defect Chip Detection in Semiconductor Manufacturing: A Wafer Spatial Context-Assisted Deep Learning Approach

Young Mok Bae, POSTECH, Pohang, Korea, Republic of

tbd

23 - A fast nonparametric process monitoring scheme in fully decentralized systems

Jiahui Zhang, University of Wisconsin-Madison, Madison, WI, United States

tbd

24 - Joint Model for Multi-Type Failure Event Prediction from Multi-Sensor TIME Series and Survival Data

Sina Aghaee Dabaghan Fard, Texas A&M University, College Station, TX, United States

TBD

25 - Point Cloud Representation Learning via Graph Signal Processing for Quality Assurance of Additive Manufacturing

Yujing Yang, The University of Texas at Arlington, Arlington, TX, United States

tbd

26 - Automated Surface Patch Extraction for 3D Printing Qualification

Weizhi Lin, University of Southern California, Los Angeles, CA, United States

tbd

27 - Uni-3DAD: GAN-Inversion Aided Universal 3D Anomaly Detection on Model-free Products

Jiayu Liu, Rensselaer Polytechnic Institute, Green Island, NY, United States

tbd

28 - Knowledge Distillation-based Incremental Learning with Integration of Transformer and Its Applications in Online Process Monitoring

Boris Oskolkov, Oklahoma State University, Stillwater, OK, United States

tbd

Sunday, October 20, 4:00 PM - 5:15 PM

SE01

Summit - 320

Platform Operations

Invited Session

Service Science

Chair: Bharadwaj Kadiyala, University of Utah, RIVERTON, United States

1 - Relaxation Methods for Hierarchical Time Series Projection

Alkiviadis Mertzios, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgia Perakis, Wei Sun, Yada Zhu

We consider the problem of Hierarchical Time Series (HTS) prediction to predict accurately time series that follow a hierarchical pattern i.e., the nodes are embedded in a tree graph and values of children nodes sum to the values of their parent. We introduce simple approaches to this problem that rely on linear regression, as well as more complicated methods based on projecting the predictions computed by linear regression. To calculate this projection we need to solve a non-convex optimization problem which is NP-hard. We use the Alternating Direction Method of Multipliers (ADMM) algorithm as a heuristic approach. This projection approach can be extended to any model that produces predictions for univariate time series, such as Random Forests or Neural Networks. We demonstrate the performance of our proposed methods to various applications including demand prediction.

2 - Food-Delivery Platforms: Near-Optimal Policies for Capacity Sizing, Order-Batching and Routing

Yang Bo, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, Milind Dawande, Ganesh Janakiraman

We study the one-time capacity sizing and infinite-horizon real-time orders' batching and routing problem for a food delivery platform. The objective is to minimize the long-run average cost incurred per unit time, where the cost includes wages to servers plus the delay penalty cost. We characterize the fundamental trade-off between spatial economies of scale and orders' waiting time, and leverage this relationship to establish a lower bound on the cost under any policy for the platform. We then identify a simple, near optimal policy whose performance gap with respect to this lower bound vanishes in a meaningful asymptotic regime.

3 - Optimal Design for Revenue-Based Financing

Eryn Juan He, University of Utah, Salt Lake City, UT, United States, Joel Goh

Small businesses are a vital part of the American economy and workforce. However, the road to success for small business is not smooth. The second most common reason that small business fail is struggling to access capital. Revenue-based financing emerges as an alternative to allow small business to get financing without giving up equity or the burden of not flexible debt conditions. This work aims to compare this model with traditional loan model and aim to develop insights into the value of revenue-based financing as well as its optimal design.

4 - Extracting Efficiency from Chaos: Rider Behavior, Design and Performance of Dockless Bike Sharing Systems

Tunay Tunca, Robert H. Smith School of Business, College Park, MD, United States, Huan Cao, Jianfeng Lin, Weiming Zhu

We empirically explore rider incentives and efficiency of dockless bike sharing systems. Using transactional data from a major Chinese bike sharing company, we estimate the usage drivers for the system. We also study measures to improve efficiency, and compare the effectiveness of dockless versus dock-based systems.

5 - Share or Solo? Platform Design and Social Choices in Ride-Hailing

Ming Hu, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Jianfu Wang, Hengda Wen, Zhoupeng Zhang

Ride-hailing platforms offer riders pooling services, enabling them to share rides with other riders. The introduction of shared rides can mitigate the driver shortage and reduce rider wait times, especially in rush hours, but it may compromise riders' privacy, space, and security. We investigate a queueing model in which riders decide whether to join the queue and, if joining, whether to opt for a shared or solo ride based on past experiences of the wait times of each option. In the event of choosing a shared ride, riders may need to wait for another fellow rider for carpooling. We examine the optimal prices for inducing self-interested riders to choose a joining rate and a sharing likelihood that maximizes throughput, revenue, or social welfare. The optimal joining rate and sharing likelihood under revenue maximization are consistently lower than those under social welfare maximization and volume maximization. As the sharing experience improves (i.e., the potentially negative sharing externality stochastically increases) and the sharing likelihood remains constant (in alignment with observations from the Chicago ride-hailing data), the revenue-maximizing shared price rises while, surprisingly, the revenue-maximizing solo price decreases. This is because the platform needs to compensate solo riders more due to a more congested system resulting from increased ride uptake. Under social welfare maximization, the price difference between solo and shared rides is less than that under revenue maximization, yet, counterintuitively, the actual prices are not necessarily lower. We corroborate our theoretical findings with a numerical study using the Chicago ride-hailing data.

SE02

Summit - 321

Quality Control in Industry 4.0

Invited Session

Quality, Statistics and Reliability

Chair: Omar Abbaas, The University of Texas at San Antonio, San Antonio, TX, United States

Co-Chair: Sara Abu Aridah, The Pennsylvania State University, State College, PA, United States

1 - Enhancing Semiconductor Yield Prediction Through Integrated Multi-View Analysis of TIME Constraint Tunnel and Equipment Usage Data

Young-Mok Bae, POSTECH, Pohang, Korea, Republic of, Chan-Young Jang, Kwang-Jae Kim

In the field of semiconductor manufacturing, yield is a crucial metric that affects the cost of overall fabrication. Yield prediction allows engineers to estimate the final cost before initiating production, effectively reducing the project risk associated with low yield. Previously, yield prediction was constrained by isolated data sources and expert judgment, particularly in the early phases of product development. As data collection capabilities within semiconductor companies have expanded, data-driven yield prediction using equipment usage data has gained traction. However, the usefulness of the equipment usage data is limited in nature because it can only show the status of the equipment. It is important to additionally leverage domain knowledge of the engineers because the semiconductor fabrication is highly process-intensive. In particular, the chemical property changes associated with the time wafers are exposed to air are known to affect the yield. This case study presents a methodology for yield prediction that integrates Time-Constrained Tunnel (TCT) data with equipment usage data through a three-step multi-view analysis approach. To enhance the interpretability of the yield prediction model, SHapley Additive exPlanations (SHAP) are also incorporated. This multi-view analysis approach enables engineers to capture the detailed temporal dynamics and equipment status, and thereby predict the yield more accurately. A case study, which utilizes actual data from a leading semiconductor manufacturer, demonstrates the effectiveness and scalability of this multi-view analysis approach.

2 - Enhancing Reduced-Order Modeling in High-Dimensional Engineering Applications through Hybrid XGBoost

Melika Baghi, Georgia Institute of Technology , Atlanta, GA, United States, Xiao Liu, Kamran Paynabar

Addressing forward problems with high-fidelity models like Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) presents significant challenges due to their computational demands from high-dimensional spatial discretization. These models are expensive, requiring powerful solvers and extensive computational resources to achieve precise results. Additionally, their sensitivity to parameter variations complicates the accurate development of Reduced-Order Models (ROMs). Projection-based methods such as Proper Orthogonal Decomposition (POD) are typically employed to extract optimal basis modes from snapshot data for ROMs. However, these methods struggle to robustly quantify uncertainty, affecting their reliability and accuracy in practical applications.

My research explores a novel integration of XGBoost, known for its predictive accuracy and capability to handle complex relationships, with an innovative mapping technique between a multi-dimensional Euclidean space and the Grassmann Manifold, which represents low-dimensional subspaces. This mapping includes an injectivity constraint to enhance optimization, making XGBoost suitable with minor modifications. The approach involves mapping the subspaces from the Grassmann Manifold back to the Euclidean space to formulate a multivariate XGBoost regression. When introduced to a new parameter, this method projects the corresponding vector onto the Grassmann Manifold, facilitating the identification of the optimal subspace or POD basis. This process allows for a detailed understanding of how variations in settings affect the optimal representation in ROMs and quantifies the confidence in these insights, addressing limitations in traditional projection-based methods.

3 - Modeling and analysis of ambient drying process for biomass materials

Chenyu Xu, University of Georgia, Athens, GA, United States, Licheng Liang, Wuyang Chen, Zipeng Guo, Dan Li, Chi Zhou, Hongyue Sun

Traditional freeze-drying techniques, while preserving the 3D structure of biomass products, are notably time and energy-intensive. This study introduces an ambient drying approach, which not only accelerates the drying process and conserves energy, but also minimizes product deformation, surpassing the capabilities of both freeze drying and conventional ambient drying methods. We developed a drying setup equipped with an adjustable-speed fan, a camera and sensors to monitor weight, humidity, temperature, and visual changes during drying. Based on degradation analysis, we aim to accurately predict the drying process for future drying optimization and control, achieving an efficient balance between time, quality, and energy consumption.

4 - Integrating Physics Information for Material Degradation Modeling and Reliability Analysis

Ali Asgari, Wichita State University, Wichita, KS, United States, Wujun Si

In recent years, there has been a growing trend in using Machine Learning (ML) to predict the Remaining Useful Lifetime (RUL) due to its robust predictive capabilities. However, traditional ML and Artificial Neural Network (ANN) algorithms often require a large amount of data during the training phase, posing a challenge in real-world industrial environments where data collection is expensive. To overcome this challenge, we propose integrating physics knowledge with machine learning for material degradation modeling and life prediction. By incorporating physics information, the data demand for training can be significantly reduced. To validate the proposed method, we conducted multiple simulation studies and a real case study regarding material degradation and life prediction.

SE03

Summit - 322

Machine Learning for Quality Control

Invited Session

Quality, Statistics and Reliability

Chair: Juan Du, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, N/A, China, People's Republic of

1 - XXXX

2 - Efficient parallel Bayesian parameter estimation of a simulation model via surrogates-assisted multiple-try Metropolis

Hyungjin Kim, Hanyang University, Seoul, Korea, Republic of, Chuljin Park, Heeyoung Kim

In semiconductor manufacturing, metrology methods have been developed to identify geometric measures, such as line width or height, of grating structures over different wafers. Among the metrology methods, the optical critical dimension (OCD) measurement is required to efficiently solve a set of parameter estimation problems of a simulation model to identify the geometric measures over different grating structures. In this work, we consider a set of parameter estimation problems in the Bayesian approach, called parallel Bayesian parameter estimation, which is motivated by the OCD measurement. The main objective is to efficiently identify the posterior distributions of parameters of a simulation model for a set of parameter estimation problems. Existing approaches focus on solving an individual parameter estimation problem, by sampling vectors that are assumed to follow a posterior distribution of parameters, via Metropolis-Hastings. However, the computational efficiency of the existing approaches significantly depends on the selection of the proposal distribution. We develop a novel framework, namely surrogates-assisted multiple-try Metropolis, that facilitates the multiple-try Metropolis with a surrogate model. Specifically, our framework introduces the proposal distributions with a surrogate model and simultaneously samples candidate vectors that are shared by a set of parameter estimation problems. We tested our framework in numerical examples and a case study in semiconductor manufacturing. The empirical results show that the proposed framework results in a high quality of the posterior distributions of parameters concerning both bias and precision while the limited simulation budget is used.

3 - Monitoring Energy Consumption from State-Space Models with Latent Variables Using Transfer Learning

Afshin Asadi, University of Miami, Coral Gables, FL, United States, Ramin Moghaddass

Identifying factors that influence the consumption behavior of industrial equipment, such as compressors and chillers, is a significant challenge that requires a comprehensive understanding of both the dynamic behavior of the equipment and hidden variables like efficiency and degradation status. Leveraging the capabilities of Variational Recurrent Neural Networks (VRNNs), our model uncovers latent patterns, contributing to a more accurate estimation of power consumption. By feeding real-time measurements into a Seq2Seq framework, enhanced with the power of LSTM (Long Short-Term Memory) networks, the model dynamically predicts power consumption over time. This VRNN architecture effectively encodes sequences into a compact representation and then decodes them to forecast future states, offering a comprehensive view of equipment performance.
Our approach stands out by combining advanced VRNN capabilities with the sequential processing strengths of the Seq2Seq architecture, enabling precise real-time predictions of power consumption in large industrial devices. This model not only captures immediate operational data but also delves into identifying underlying trends and patterns, providing insights for proactive energy management. The integration of these techniques results in a robust tool that goes beyond traditional monitoring, offering predictive analytics and optimization strategies for complex industrial environments. Additionally, a transfer learning method can be implemented to utilize vast historical data, making the model more accurate and effective by leveraging previously learned patterns and knowledge.

SE04

Summit - 323

Statistical Machine Learning for Complex Data

Invited Session

Quality, Statistics and Reliability

Chair: Xiaoyu Chen, University at Buffalo, Buffalo, NY, 14051, United States

Co-Chair: Chenang Liu, Oklahoma State University, Stillwater, OK, 74078, United States

1 - Multitask Process Adjustment for Similar-but-non-identical Products and Processes in Cyber Manufacturing Systems

Yingyan Zeng, University of Cincinnati, Cincinnati, OH, United States, Xiaoyu Chen, Lening Wang, Ran Jin

A cybermanufacturing system aims to facilitate interconnected manufacturing processes with computational services, such as modeling and control. However, highly customized demands and heterogeneous manufacturing equipment in Cyber Manufacturing systems have posed significant challenges to the modeling and control needs of products' quality. For example, limited samples extracted from a customized product design with invariable process settings cannot meet the degree of freedom requirement of many quality-process models, hence leading to unsatisfactory modeling accuracy and control performance.
Therefore, it is important to consider the information (i.e., process settings and *in situ* process variables) shared from similar-but-non-identical products and equipment. In this research, we propose a multitask process adjustment (MPA) method to adjust process setting variables by jointly modeling setting-process and quality-process relationships. Defining a task as one product design or one manufacturing equipment, MPA method introduces variability to invariable process settings for one task by considering different settings from multiple tasks. Therefore, MPA method can adjust the process by optimizing setting parameters especially for the task without variability. The merits of the proposed method are illustrated by a feedforward control case studies in a selective laser melting (SLM) process.

2 - A Multimodal AI Framework for Motor RRBs Detection and Forecasting Among Children with ASD

Mengqi Shen, Columbia University, New York, NY, United States

TBD

3 - Distribution-in-distribution-out Regression

Xiaoyu Chen, University at Buffalo, Buffalo, NY, United States, Mengfan Fu, Yujing Huang, Xinwei Deng

Regression analysis with probability measures as predictor and response has gained increasing attention. However, extending existing univariate regression model to multiple linear regression model is challenging due to the non-flat Riemannian geometry of the Wasserstein space, hindering the definition of arithmetic operations, hence additivity is not well-defined. In this paper, a distribution-in-distribution-out regression model is defined by introducing parallel transport to enable provable commutativity and additivity of newly defined arithmetic operations in Wasserstein space. The Frechet least squares estimator is hence employed to obtain the best linear unbiased estimate via the newly established Frechet Gauss-Markov Theorem. To accommodate the practical needs, we further investigate a special case when predictors and response are all univariate Gaussian measures, resulting in a close-form solution of linear model coefficients and R-square metric. A simulation study and real case study in intra-operative cardiac output prediction are performed to evaluate the performance.

4 - Deep Distribution-in-distribution-out Neural Networks on Riemannian Manifold

Yujing Huang, University at Buffalo, Buffalo, NY, United States, Mengfan Fu, Xiaoyu Chen

Data collected from real-world systems commonly displays uncertainty, due to the randomness introduced by imperfect measurements and unknown physical processes. A natural method for addressing such uncertainty involves representing the data through probability distributions. However, existing statistical models are predominantly tailored for Euclidean data (e.g., scalar-, vector-, matrix-valued data) and thus fail to directly quantify relationships among probability distributions residing in non-Euclidean spaces.
Few studies have examined univariate regression models that regress a distribution-valued response on a distribution-valued predictor. Extension of these models to multiple regression is impeded by the lack of well-defined arithmetic operations, such as addition and scalar multiplication. A novel approach, the distribution-in-distribution-out (DIDO) regression model, establishes a linear additive framework on Riemannian manifolds, specifically Wasserstein spaces, enabling the construction of multilinear regression models for probability densities. Nevertheless, considering nonlinear regression model on a Riemannian manifold remains unclear, owing to undefined higher-order effects in the DIDO framework.
To model nonlinear relationships, this work defines a neural network structure based on the DIDO regression model on a Riemannian manifold. This research presents the first model capable of effectively handling distribution-valued data, expanding DIDO from single output to multiple outputs with a nonlinear structure. Additionally, our neural network architecture more accurately approximates underlying nonlinear relationships, thus enabling the modeling and inference of nonlinear system dynamics. Quantifying the nonlinear relationships among probability distributions is expected to broadly transform the current analytical tools by directly considering uncertainties in individual samples.

SE05

Summit - 324

DEI Best Student Paper Award II

Award Session

Diversity, Equity, and Inclusion

Chair: Zeyu Liu, West Virginia University, Morgantown, WV, United States

Co-Chair: Albert Berahas, University of Michigan, Ann Arbor, MI, United States

1 - Fair Fares for Vehicle Sharing Systems

Hyemi Kim, Columbia University, New York, NY, United States, Adam Elmachtoub

Vehicle sharing systems, like those for bicycles, scooters, and cars, have become essential transportation options, with fares often set by companies through mobile apps. This can result in users in different locations paying varying prices and experiencing unequal access. Platforms and regulatory bodies should consider how to set prices fairly to minimize the inequalities experienced by users across locations, for which we provide a framework and insights in this work. We consider two notions of fairness corresponding to price and access. Price fairness is a measure of how close the prices at two locations are to one another. Access fairness compares the fraction of demand at each location that has access to the system, where access is a product of affordability and vehicle availability. We analyze the impact of imposing these fairness measures on the revenue, consumer surplus, and social welfare using a stylized model with two locations. Under price fairness, we analytically identify regimes where consumer surplus at both locations can increase or can both decrease. We show that access fairness decreases consumer surplus at both locations, making all parties worse off. To address the non-convexity in generalized networks with multiple locations and vehicles, we design a sliding window method for price fairness and employ a convex relaxation technique achieving asymptotic optimality for access fairness. Lastly, a case study using real-word data of a vehicle sharing system operating in New York City shows that a significant gain in fairness can be achieved with a relatively small loss in revenue

2 - Foundations for Equitable Transit Networks

Sophie Pavia, Vanderbilt University, Nashville, TN, United States, abhishek dubey, Ayan Mukhopadhyay, J. Carlos Martinez, Philip Pugliese, Samitha Samaranayake

Public transit is vital for accessing essential services such as education, healthcare, and employment. While transit accessibility is important in general, some sections of the population depend critically on it. However, existing paradigms for public transit design neglect equity as a design objective, which hampers systemic progress toward equitable transit infrastructure. We present a mathematical formulation for transit network design that explicitly considers different constructs of equity and welfare. Our formulation is a mixed-integer linear program (MILP) based on a (linearized) piece-wise linear utility function that quantifies the utility a passenger reaps from the installed network compared to using personal vehicles. We also present two algorithmic approaches for equity-aware transit network design. To inform our model, we conduct a community survey of 673 residents from the city of Chattanooga, USA. We study the interaction between network design and different concepts of equity and showcase trade-offs based on real-world data from two metropolitan cities.

3 - Potty Parity: Process Flexibility via Unisex Restroom

Setareh Farajollahzadeh, University of Toronto, Toronto, ON, Canada, Ming Hu

We address unequal restroom access for women and LGBTQ+ individuals, known as the ``potty parity" problem. We propose a utility model where users consider gender identity, wait time, and safety concerns when choosing restrooms. We evaluate different layouts' efficiency by total utilities (totalitarian principle) and assess their fairness using minimum utility gain (Rawlsian fairness) and the gap between maximum and minimum gains (distributive fairness). While it may initially seem intuitive to assume that converting all restrooms to unisex facilities would be efficient due to the pooling of servers and increased flexibility and fairness due to all users standing in the same line, our findings demonstrate that this design can be neither efficient nor fair. In contrast, we show that converting some of the men's restrooms to unisex facilities can enhance both efficiency and fairness of access. This highlights that a moderate level of flexibility can surpass a fully flexible system. Moreover, conventional wisdom suggests that removing a unit of restroom from the men's room would negatively impact users from the men's side. However, our analysis reveals a counterintuitive result that such a change can lead to a Pareto improvement, benefiting all users involved. We also analytically explore additional benefits of unisex restrooms under different user behaviors and situations and present numerical results to support our findings. Moreover, we provide guidelines for inclusive renovation plans of restroom layouts based on the comfort levels of different population groups regarding sharing restrooms with users of other genders.

SE06

Summit - 325

Location Analysis for Architectural Planning

Invited Session

Location Analysis

Chair: Kaori Isawa, Institute of Industrial Science, The University of Tokyo, Meguro, Japan

Co-Chair: Yudai Honma, The University of Tokyo, Meguro-ku, 153-8505, Japan

1 - Possibility of energy saving management through demolition and renovation of existing building stock

Kaori Isawa, The University of Tokyo, Meguro, Japan, Yudai Honma

In recent years, there has been a growing emphasis on adhering to the Sustainable Development Goals (SDGs); if we adhere to the SDGs to reduce CO2 emissions, we should consider replacing old, low-performance building stocks with new, high-performance ones. However, if we also follow the SDGs to utilize limited resources, we should continue using the existing building stocks. This dilemma presents a challenge in urban planning, as it involves balancing the objectives of demolishing and utilizing existing building stocks, namely, promoting new construction and encouraging renovation efforts. Therefore, in this study, we aim to address this conflict by proposing a framework for sustainable urban development, focusing on energy-efficient practices supported by quantitative evidence. Specifically, we seek to explore strategies integrating new construction and renovation initiatives to achieve optimal urban metabolism. Concerned with the current situation in Japan, which is a strong preference for new construction, we analyze the optimal effective subsidy investment pattern on Energy saving policy to design systems that make renovation a strong and attractive option. Specifically, we analyze it, focusing on the relationship between annual income and household distribution. In addition, we pay attention to the public's aversion to second-hand property and the mismatch between the utility of the public, who prefer to reduce construction costs without worrying about CO2 emissions and the government, who prefer to reduce CO2 emissions, which enables more real-world results to obtain.

2 - Proposal and Mathematical Examination of a Trajectory Estimation Model from Sparse and Noisy Location Data

Junya Maruyama, The University of Tokyo, Meguro, Japan, Yudai Honma

It is important to observe the actual behavior of users in order to analyze the utilization of buildings and the value of the space. On the other hand, indoor positioning within buildings, compared to outdoor positioning using GPS, presents technical difficulties due to the details of the space and the effects of obstacles. Regarding this problem, the trajectory estimation model, which is a mathematical model for estimation actual trajectories based on observation data, has been evolving. Existing trajectory estimation techniques have pursued the technology to observe as much observation data as possible and to observe very detailed trajectories. However, introducing such observation modules is often difficult, and the accuracy of the trajectory is often excessive. Therefore, this study naturally constructs a model to observe trajectories with appropriate detail using sparse and noisy location data and redefines the model as an estimation algorithm using the Hidden Markov Model (HMM) commonly used in existing research. The goal of this trajectory estimation model is to extract trajectories at the meter level from room-level observation data. The initial trajectory estimation model is formulated as a natural and intuitive mathematical model that minimizes the sum of the squares of the distances between stay points and location information, as well as the moving distance. This formulation can be redefined as an algorithm using the HMM. By performing this redefinition, it becomes possible to examine the characteristics and advantages of the proposed model, which will provide useful information for future studies on the accuracy of trajectory estimation.

3 - Year-Round Shadow Analysis for Pedestrian Space in Vertical Urban Environments

Hiroko Watanabe, Institute of Industrial Science, The University of Tokyo, Meguro, Japan, Yudai Honma

In recent years, as urban areas have increasingly developed vertically, there has been a growing trend toward creating walkable cities that maintain a human scale. However, with the intensifying effects of global warming, summer heat has become more severe, with last summer's unprecedented heatwaves still fresh in memory. Despite these harsh outdoor conditions, people still desire to walk through the city and enjoy its attractions. In Japan, people prefer walking in sunny spots during winter and shaded areas during summer, indicating a need for comfortable pedestrian spaces. This study aims to understand the shadows cast by buildings throughout the year, which is crucial for designing pleasant pedestrian environments. In highly vertical urban areas, the extensive shadows from tall buildings significantly impact pedestrian spaces. Given that the preferred shadow conditions for pedestrians vary with the seasons, it is essential to monitor building shadows throughout the year. Therefore, this research conducts large-scale and detailed spatiotemporal shadow simulations to capture these dynamics. By understanding the annual variations in shadows, we can demonstrate their impact on pedestrian spaces, ultimately contributing to the design of more comfortable urban environments.

4 - A Study on Spatial Cognition of Visitors Using Mathematical Analysis of Visitor's Movement Within An Art Museum Gallery

Yunqing Hu, Institute of Industrial Science, The University of Tokyo, Meguro, Japan, Junya Maruyama, Kaori Isawa, Yudai Honma

In China, art museums are gaining more attention nowadays but visiting them is still not a popular lifestyle. With the emergence of open-plan galleries with non-fixed routes in the beginning of 21st century, the necessity of the then trendy large-scale central space and the traditional sequential art viewing experience is challenged. In this research, we aim to understand the navigation patterns of visitors and through which to identify their spatial cognition in open-plan gallery spaces. First, to understand the navigation patterns of visitors, a random Delaunay network (rDn) is applied to the target plan, and the hidden Markov model is introduced, in which the hidden state being the location of visitor, and the observation state as the artwork being viewed. Given various artwork viewing sequences, we are able to discover the best hidden state sequences of transition from locations to locations, thus obtain possible viewing routes within the target art gallery. Second, to identify visitor's spatial cognition with in the target art gallery, the resulted viewing routes is clustered and analyzed. The goal of this research is to develop a methodological reference for future art museum design, to create more relaxed and welcoming art viewing and interacting environment, and to assist with the promotion of art museum culture and lifestyle to the society from an architectural point of view.

5 - Mathematical Consideration on the Method for Enumerating Spaces "Without Omissions and With Omnipresence" for the Algorithm of Automatic Generation of Architectural Spaces

Keito Nakagawa, The University of Tokyo, Setagaya-city, Tokyo, Japan, Yudai Honma

Previous studies have focused on the "form" of the architectural space and conducted mathematical studies of methods to analyze it. One such method is to consider the inner-convex space. Convex space has the property that any people in it can interact with each other, and it can be one of the indicators to estimate where in the architectural space the interaction of people is promoted. However, the proposed algorithm for enumerating inner-convex spaces does not enumerate them in a spatially "without omissions and with omnipresence" manner. In this study, we assume that enumerating a space "without omissions and with omnipresence" is an important factor in human consideration of the inner space. Enumerating the space "without omissions and with omnipresence" is also useful in the construction of automatic design algorithms. It is realistic to use an automatic design algorithm in which several plausible patterns are presented and compared with each other, and the final plan is determined by human consideration. In order to enumerate spaces "without omissions and with omnipresence," we first focus on the generation probability of inner spaces. The existence of boundaries causes variations in the probability of its creation. We will examine this using geometrical methods and mathematical optimization algorithms, and clarify the characteristics of inner spaces. Next, we clarify the conditions necessary to enumerate spaces "without omissions and with omnipresence" from the viewpoints of both computer experiments and mathematical theory.

SE07

Summit - 327

Innovations in Logistics and Transportation Networks

Flash Session

Contributed

Chair: Amin Keramati, Widener University, New York, 19082, United States

1 - Introducing a Web Application Leveraged by An Optimized Proximity Algorithm to Improve Logistics Network Efficiency and Reduce Travel TIMES

Amin Keramati, Widener University, Chester, PA, United States

This study introduces a web application powered by a proposed Optimizing Proximity Algorithm designed to enhance logistics network efficiency by improving travel time. The application leverages advanced spatial analysis and data-driven approaches to identify the nearest facilities—such as restaurants, retailers, storage units, and warehouses—based on user-defined criteria in the search function. It performs routing and travel duration analyses to minimize travel distances and optimize delivery speeds, taking into account various business factors like transportation modes, departure times, and operating hours.

The presentation will highlight how the application and its algorithm adeptly process both real-time and historical logistics data, considering elements like traffic patterns and geographical constraints, utilizing tools like the Google API and Google Places library. By offering interactive maps and outputs, the application supports more informed decision-making in logistics management. This leads to optimized resource and location utilization, cost reductions, and a significant decrease in the environmental footprint of logistics activities.

2 - Density Tensor Factorization for Sediment Source Analysis

Nicholas Richardson, University of British Columbia, Vancouver, BC, Canada, Naomi Graham, Michael Friedlander, Joel Saylor

Detrital zircon grains carry history about the formation of geological structures and sources of precious minerals. DZ Grainalayser is a new tool we develop for identifying sediment sources from mixtures of detrital zircon grains with arbitrary number of features. We test the model's effectiveness at using constrained tensor factorization to demix high dimensional distributions on real world sediment data. Finally, we show its portability to diverse tasks like identifying cell types in spatial transcriptomics.

3 - A Neural Optimization Machine Approach for the Fixed-Price Game Among E-hailing Drivers

Hanwen Dai, Tsinghua University, Beijing, China, People's Republic of

This study introduces a novel approach to optimize the allocation of fixed-price drivers in a dynamic ride-sourcing environment. Initially, a forecasting neural network is trained to predict the joint order matching probabilities for various driver types. Leveraging the forecasting model, a Neural Optimization Machine (NOM) is devised to determine the optimal number of fixed-price drivers.

Ride-sourcing market involves diverse drivers with varying characteristics such as vehicle types and activeness. This heterogeneity poses a significant challenge in accurately forecasting order matching probabilities across different driver categories. Employing neural networks, which excel in capturing intricate nonlinear relationships, we analyze real-world operational data from a Chinese ride-sourcing company.

Drivers' decision to accept fixed-price orders impacts the supply structure. With our fitted model, we investigate the ideal proportion of fixed-price drivers for each driver type, aiming to either maximize total service reward or ensure fairness among drivers. Yet, neural networks are often perceived as opaque due to their layered structure and nonlinear activation function. To address this challenge, we introduce a NOM that operates on top of the forecasting model. It incorporates an additional objective construction layer to translate predicted matching probabilities into a target objective. Furthermore, a decision layer, powered by a customized softmax activation function, categorizes drivers into fixed-price and non-fixed-price groups. During the optimization process, the network's backpropagation algorithm updates only the parameters of the decision layer while keeping other layers fixed. Once the objective function converges, the updated weights of the decision layer yield the optimal driver allocation strategy.

4 - Chance Constrained Problem Based On Compound Poisson-gamma Distribution

LANQING DU, Drexel University, Philadelphia, PA, United States, Jinwook Lee

This study combines a compound probability distribution with a chance-constrained model to enhance its capacity in representing various intricate business operations. Specifically, we focus on the numerical techniques used to manage uncertainties, introducing an algorithm that effectively implements multivariate p-quantile (p-level) efficient points under independent and identically distributed (i.i.d.) conditions. To elaborate on the numerical results, our algorithm is applied within a framework of finite-period equipment portfolio problem.

5 - Interpretable Tabular Reinforcement Learning Applied to Revenue Management Problems

Adam Page, Lancaster University, Lancaster, United Kingdom, Anna-Lena Sachs, Christopher Kirkbride, James Grant

Motivated by the challenge of pricing pre-bookable airport car parking, we consider revenue management problems with non-stationary demand. Customers differ by each having their own willingness to pay, buying preferences, lead time and length of stay. Instead of using classical dynamic programming approaches which suffer from the curse of dimensionality, we propose tabular reinforcement learning methods to find an interpretable solution while reducing the effect of added dimensionalities. These methods can be implemented online which allows an agent to react to changes in customer buying patterns. We show the effectiveness of algorithms such as Q-learning with cross-learning to tackle these challenges.

SE08

Summit - 328

Operations Planning to Ensure Food Security

Contributed Session

Contributed

Chair: Clara Novoa, Texas State University, 601 University Dr, San Marcos, TX, United States

1 - A Conceptual Framework for Decision Support Systems in Food Safety Investigations

Sandra Rudeloff, Kühne Logistics University, Hamburg, Germany, Hanno Friedrich

Unsafe food results in approximately 600 million cases of foodborne illnesses and 420,000 deaths worldwide, posing a significant global health and economic challenge. Despite advancements in food safety practices and investigations, these outbreaks continue to occur and often remain unsolved or are resolved too late to prevent further illnesses.

Traditional methods for investigating outbreaks, predominantly relying on patient interviews, are hampered by recall bias and substantial time demands. To address these limitations, emerging methods that leverage previously untapped data sources, such as retail data or food supply networks, show promise. However, the practical implementation of these methods is often restricted by a lack of awareness or understanding, as well as constraints related to data availability. This creates a significant gap in the comprehensive overview and effective use of both traditional and new methods, preventing investigators from fully utilizing all available options to resolve ongoing outbreaks.

To bridge this gap, we propose a conceptual framework designed to guide the development of decision support systems for food safety investigations. This framework systematically outlines potential investigative pathways and pinpoints relevant data sources for identifying contaminated food items and their origins. It assesses how data quality and availability may vary based on geographical location and outbreak contexts. Additionally, by demonstrating effective utilization of pre-collected data, the framework promotes their systematic collection, thereby enhancing preparedness for future outbreaks. Our framework broadens the investigative toolbox for food safety by incorporating alternative methods and data, offering more diverse options to effectively tackle foodborne outbreaks.

2 - Measuring Food Waste Impact of Perishable Inventory Management Solutions

Clement Micol, Afresh Technologies, Brooklyn, NY, United States, Marc Goessling, Aaron Stern, Mark Velednitsky, Carl Vogel

As sustainability goals are receiving ever-increasing attention, grocery corporations are incentivized and expected to track their food waste. The common approach in the industry is to compute some notion of missing product units, referred to as "shrink." However, accurately calculating food waste at the store level over a given time period is a difficult task in the presence of imperfect shipment, sales, and inventory data.

In this talk, we discuss three different food waste metrics, the practical challenges we encountered when using them and the corresponding solution approaches we explored. One source of large measurement noise when tracking food waste is the inaccuracy of ending inventory counts, caused for example by typos or wrong item lookups. Item-level errors can be combated through data quality checks and data filtering. We also discovered that subtle differences or changes in truck delivery schedules can cause an outsized effect on food waste metrics for a given accounting period. As a solution in synthetic control analyses, we propose a post-hoc alignment procedure on delivery schedules.

Lastly, we discuss the benefits and difficulties of food waste measurement through scan-outs, which are unit counts of expired items provided by grocery store associates.

3 - Preventing Food loss and Food waste in Food Supply/Value Chains

Ramesh Bollapragada, San Francisco State University, San Francisco, CA, United States

This project motivated by the Food loss and Food waste that is prevalent in current Food supply chains. Food loss occurs before the consumer and supermarket level that includes crop loss due to pests, farming techniques, insufficient storage space, and obsolescence due to logistics bottlenecks in the supply chain. Food waste happens at the end consumer level (e.g. restaurants, households discarding food due to lower consumption than the amount produced, and obsolescence arising from poor inventory management). The Food loss and waste in the world is estimated at 30-40% (i.e. more than a third of the food made is wasted). The estimated loss and wastage for fruits and vegetables is even higher (40-50%). These numbers amount to a loss of over $2.5 trillion USD per year globally (with half a trillion USD in United States alone), while over a billion people globally are suffering from malnutrition and unavailability of food.

We present a framework that outlines the above issues in developed and developing countries, identifying the areas of improvement to make a significant impact in savings in the above loss.

4 - Identifying Critical Factors Affecting Inventory Costs and Optimizing the Pick-up and Delivery of Donated food at the Central Texas Food Bank

Clara Novoa, Texas State University, San Marcos, TX, United States, Jakir Hassan

We present the results of experimenting with a multi-period multi-product inventory model developed for assisting the Central Texas Food Bank with their food procurement plans. The factors (i.e., model parameters) that affect more significantly the model costs are identified.

We also present a mixed-integer programming model to optimize the pick up and delivery of food procured and donated considering an heterogeneous fleet of vehicles. The routing model provides a cost-efficient plan for designing the daily routes.

5 - Heterogeneity of willingness to accept for using reusable cups based on framing effect: A causal forest approach

Sung-Eun Chang, Korea University, Seongnam-si, Korea, Republic of, JongRoul Woo

Worldwide, 139 million tons of single-use plastic waste was generated in 2021. Accordingly, governments and commercial facilities implement a variety of efforts to encourage consumers to continue their sustainable behavior. Among them are incentive programs that offer monetary benefits when customers bring their own reusable containers/cups. Persuasive communications could be a key part of increasing consumer acceptance of these programs. Relatedly, framing messages in terms of gains and losses has been extensively studied and has been shown to be effective in many cases. However, the impact of this framing can also be heterogeneous depending on the consumer. In this study, we analyze the treatment effects of framing on the intention to use reusable cups for individuals. Causal Forest, a machine learning methodology, can effectively identify the heterogeneity of treatment effects between subgroups in the data and capture complex non-linear relationships and interactions. These findings provide policy implications for effective incentive programs for the use of reusable cups.

SE09

Summit - 329

Operations and Market Design for Education

Invited Session

Auctions and Market Design

Chair: Faidra Monachou, Yale University, New Haven, CT, United States

1 - Optimal Cohort Partitions

Faidra Monachou, Yale University, New Haven, CT, United States, Sofoklis Goulas

The optimal cohort partition problem involves the partition of a heterogeneous population of individuals into distinct groups to optimize the objective of a social planner. Unifying results from optimization and majorization theory, we characterize the structure of optimal cohort partitions across various objectives and applications. Our primary focus is on a common educational application: how to optimally partition students into classes in the presence of peer effects. We employ the workhorse linear-in-means (LIM) model from the empirical literature on peer effects in education and study two different behavioral microfoundations: the LIM spillover model and the LIM conformist model. We theoretically characterize the optimal partition for various objectives---performance, welfare, diversity, inequality---and find that, within the same behavioral model, trade-offs are not always inevitable. We derive several policy-relevant insights. First, in both behavioral models, we show that as the planner's highest priority shifts from the top-performing class to the least-performing class, the optimal partition becomes progressively less assortative. Second, we show that, while both behavioral models agree on the optimal partition policy for performance, diversity, and inequality, they disagree on the welfare-optimal policy. Third, we find that the commonly used uniform partition, despite achieving lower inequality than any integral partition, does not achieve perfect equality. We illustrate these theoretical findings in simulation exercises with student data from a novel dataset of public high schools in Greece. Finally, we showcase how our theoretical framework extends beyond educational contexts by characterizing the optimal assignment of experts to teams and the design of Tullock contests.

2 - Stable Matching with Contingent Priorities

Ignacio Rios, The University of Texas at Dallas, Richardson, TX, United States, Federico Bobbio, Margarida Carvalho, Alfredo Torrico

We study stable matching problems under contingent priorities, whereby the clearinghouse prioritizes some agents based on the allocation of others. Using school choice as a motivating example, we first introduce a stylized model of a many-to-one matching market where the clearinghouse aims to prioritize applicants with siblings assigned to the same school and match them together. We provide a series of guidelines to implement these contingent priorities and introduce two novel concepts of stability that account for them. We study some properties of the corresponding mechanisms, including the existence of a stable assignment under contingent priorities, its incentive properties, and the complexity of finding one if it exists. Moreover, we provide mathematical programming formulations to find such stable assignments whenever they exist. Finally, using data from the Chilean school choice system, we show that our framework can significantly increase the number of siblings assigned together while having no large effect on students without siblings.

3 - Rethinking the Role of Teacher Evaluation: An Analysis of Its Impact on Quality, Workforce Dynamics, and Student Outcomes

Hemanshu Das, Yale University, New Haven, CT, United States, Edieal Pinker

In the realm of public education, the quality of teachers is an influential factor that shapes student outcomes and, subsequently, their future prospects. The study delves into the impact of teacher evaluation programs, on teacher quality, student outcomes, budget considerations, workforce dynamics, and hiring requirements. A case study of the District of Columbia Public Schools (DCPS) illustrate the diverse outcomes of evaluation systems. DCPS's IMPACT evaluation system yields 0.46 standard deviation increase in teacher quality, leading to incremental classroom incomes of $67,403 but comes at an additional cost of $16,992 per teacher annually. The study elucidates the multifaceted considerations involved in teacher evaluation systems, emphasizing the need to weigh factors such as retention probabilities, availability of proficient teachers, correlation between evaluation measures, and budgetary implications. By offering quantitative insights into these dynamics, we hope to equip policymakers and educators with guidance for enhancing teacher quality and elevating student outcomes within public education systems.

4 - 30 Million Canvas Grading Records Reveal Widespread Sequential Bias and System-Induced Surname Initial Disparity

Jiaxin Pei, University of Michigan, Ann Arbor, MI, United States, Zhihan Wang, Jun Li

The widespread adoption of learning management systems in educational institutions has yielded numerous benefits for teaching staff but also introduced the risk of unequal treatment towards students. We present an analysis of over 30 million Canvas grading records from a large public university, revealing a significant bias in sequential grading tasks. We find that assignments graded later in the sequence tend to (1) receive lower grades, (2) receive comments that are notably more negative and less polite, and (3) exhibit lower grading quality measured by post-grade complaints from students.

Furthermore, we show that the system design of Canvas, which pre-orders submissions by student surnames, transforms the sequential bias into a significant disadvantage for students with alphabetically lower-ranked surname initials. These students consistently receive lower grades, more negative and impolite comments, and raise more post-grade complaints as a result of their disadvantaged position in the grading sequence. This surname initial disparity is observed across a wide range of subjects, and is more prominent in social science and humanities as compared to engineering, science and medicine. The assignment-level surname disparity aggregates to a course-level surname disparity of students' GPA and can potentially lead to inequitable job opportunities. For platforms and education institutions, the system-induced surname grading disparity can be mitigated by randomizing student submissions in grading tasks. Education institutions should keep the workload of graders at a reasonable level to reduce fatigue and/or have multiple graders as a cross validation to enhance grading quality.

SE10

Summit - 330

Job Market Candidates Session Focusing on Social Impact Operations

Invited Session

Auctions and Market Design

Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

1 - You Can't Shake Your Past -- Prior Ratings, Gender and Instructor Outcomes in Higher Ed

Tong Wang, London Business School, London, United Kingdom, Kamalini Ramdas, Monika Heller

Instructor ratings facilitate students' course choice decisions. While theory suggests that quality signals can introduce bias, the impact of prior teaching ratings remains unstudied in Higher Ed – where there is evidence of gender bias in traditional classrooms. We ran controlled experiments to examine these sources of bias in online Higher Ed delivery. Experimental findings indicate that low prior ratings significantly diminish the likelihood of positive evaluations, particularly for novice learners and for instructor-focused outcomes. Also, gender bias persists online. Female learners exhibit bias against male instructors, while male learners exhibit bias against female instructors, with effects concentrated on conditions where learners face high cognitive load due to time pressure. The bias against male instructors is more pronounced among older or less educated learners, whereas the bias against female instructors is more pronounced among younger and more educated learners. Furthermore, we find evidence of a strong interaction between prior ratings and instructor gender for male learners. Without time pressure, male learners exposed to low prior ratings correct the unfairly low ratings for male rather than female instructors. Under high time pressure, male learners exposed to high prior ratings increase their ratings for male instructors, exhibiting confirmation bias. Our research highlights the deleterious consequences of the widespread practice of sharing instructor ratings with learners. We also provide evidence of gender inequity under different prior ratings and time pressure conditions, highlighting potential levers to mitigate such bias.

2 - Designing Algorithmic Recommendations to Achieve Human-AI Complementarity

Bryce Mclaughlin, University of Pennsylvania, Philadelphia, PA, United States, Jann Spiess

Algorithms frequently assist, rather than replace, human decision-makers. However, the design and analysis of algorithms often focus on predicting outcomes and do not explicitly model their effect on human decisions. This discrepancy between the design and role of algorithmic assistants becomes of particular concern in light of empirical evidence that suggests that algorithmic assistants again and again fail to improve human decisions. In this article, we formalize the design of recommendation algorithms that assist human decision-makers without making restrictive ex-ante assumptions about how recommendations affect decisions. We formulate an algorithmic-design problem that leverages the potential-outcomes framework from causal inference to model the effect of recommendations on a human decision-maker's binary treatment choice. Within this model, we introduce a monotonicity assumption that leads to an intuitive classification of human responses to the algorithm. Under this monotonicity assumption, we can express the human's response to algorithmic recommendations in terms of their compliance with the algorithm and the decision they would take if the algorithm sends no recommendation. We showcase the utility of our framework using an online experiment that simulates a hiring task. We argue that our approach explains the relative performance of different recommendation algorithms in the experiment, and can help design solutions that realize human-AI complementarity.

3 - Improving compliance to enhance outcomes in algorithmic refugee assignment

Anand Kalvit, Stanford University, Stanford, CA, United States, Yonathan Gur, Jens Hainmüller

The assignment of refugees to resettlement locations within a host country presents complex challenges, particularly given the limited capacities at these locations. Traditionally, placements are decided by case officers based on their experience and expertise, which may not align with the recommendations of data-driven algorithms designed to optimize long-term outcomes while considering capacity constraints. This misalignment can lead to inefficiencies and suboptimal matches. To address this issue, we propose an innovative approach that incorporates future-accountability into the decision-making process by linking a case officer's utility to both immediate and future allocations. This approach increases compliance with algorithmic recommendations, thereby enhancing the overall effectiveness and efficiency of the resettlement system. We demonstrate the benefits of our approach through numerical experiments, showing significant improvements in match quality and system performance.

4 - Childcare Shortage and Talent Retention in the Knowledge Industries

Zhihan (Helen) Wang, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Damian Beil, Jun Li

Achieving an equitable workplace goes beyond ensuring unbiased treatment at work; it also involves addressing the hidden burdens employees face in managing household responsibilities, where childcare is a critical component. The shortage of reliable childcare services can hinder the career progression of young professionals, resulting in higher turnover rates and a more volatile workforce. This challenge is especially pronounced in knowledge-driven sectors like finance, technology, and professional services, where high turnover and frequent handoffs create significant inefficiencies due to the costly nature of knowledge transfer. Leveraging a comprehensive national job profile data from 1997 to 2019, we investigate how the turnover rates of workers in knowledge industries are influenced by the availability of childcare services nearby their working sites. This research offers practical implications for both social planners and employers, emphasizing the need for accessible, conveniently located childcare services to enhance the well-being and productivity of working parents.

SE11

Summit - 331

Network Analytics at Amazon Last Mile Logistics

Panel Session

Telecommunications and Network Analytics

Co-Chair: Natarajan Gautam, Syracuse University, Syracuse, NY, United States

1 - Moderator Panelist

Natarajan Gautam, Syracuse University, Syracuse, NY, United States

2 - Panelist

Abhilasha Katariya, Amazon, Seattle, WA, United States

3 - Panelist

Dipal Gupta, Amazon, Seattle, WA, United States

4 - Panelist

Chinmoy Mohapatra, Amazon, Lake Stevens, WA, United States

5 - Panelist

Gokce Kahvecioglu, Amazon, Bothel, WA, United States

SE12

Summit - 332

Vehicle Routing and Last Mile Delivery

Flash Session

Flash

Chair: Hyeseon Lee, 1400 Townsend Dr, 49930

1 - Evaluating Temporally-Consolidated Last mile Delivery Model: Integrating Per-order and Subscription Pricing

Behnaz Naeimian, York University, Toronto, ON, Canada, Opher Baron, Mehdi Nourinejad, Peter Park

As same-day delivery continues to appeal for its convenience and efficiency, its substantial operational costs remain a barrier to broader adoption. This study proposes a "Temporally-Consolidated" (TC) delivery model, specifically designed for recurring purchases such as groceries, which schedules deliveries on predetermined days to minimize the transportation costs linked to sporadic same-day deliveries. This model depends on customer flexibility to adhere to predefined delivery schedules. We explore two pricing strategies within the TC framework: per-order charges and a subscription model. Our analysis assesses customer utility for each pricing strategy, including comparisons to traditional same-day delivery. By employing a profit model that incorporates these utilities, we assess the impact of these strategies on overall profitability. Results indicate that a mixed approach, combining both per-order and subscription plans, maximizes profitability by accommodating varying customer buying patterns: infrequent purchasers prefer per-order plans, while frequent buyers opt for subscriptions. This study offers significant insights into the viability of TC delivery, underscoring its potential to enhance profitability and meet the needs of diverse consumer segments.

Keywords: Last mile delivery, temporally-consolidated delivery, customers' utility, frequent and infrequent customers, subscription pricing, per-order pricing, transportation costs, profit model.

2 - Key Factors of Service Time for Urban Deliveries in the U.S.

Oriana Calderon, Rensselaer Polytechnic Institute, Troy, NY, United States, Andres Regal Ludowieg, Jose Holguin-Veras

The significant growth of e-commerce in recent years has led to a higher number of packages delivered to households and an increase in truck stops across the United States, which has, in turn, affected the service time for urban deliveries. In light of this, this research aims to identify the key socio-economic, spatial, and package attributes that determine service time—defined as the time between when the truck driver parks and when the package is delivered—for urban deliveries. This research utilizes the publicly available dataset from the 2021 Amazon Last Mile Routing Research Challenge and focuses on five major metropolitan areas: Austin, Boston, Chicago, Seattle, and Los Angeles. An interaction index was used to identify the economic center of the cities. Through econometric modeling, the research identifies the most significant factors that determine service time for urban deliveries in these areas. The results reveal interesting commonalities across the metropolitan areas considered, including shorter service times associated with larger household sizes, shorter service times with increased distance from the delivery location to the economic center of the cities, and longer service times for larger package sizes. Understanding these key factors is essential for optimizing delivery processes, reducing delivery times, and enhancing overall efficiency in urban logistics, thereby promoting more sustainable delivery systems.

3 - Add to Order in Contingent Free Shipping

Yuxuan Tang, National University of Singapore, Singapore, Singapore, Xiaobo Li, Rowan Wang, Xuan Wang

This paper explores the impact of a contingent free shipping policy combined with an "add to order" option, recently implemented by online grocery and food delivery platforms such as Amazon Fresh, Instacart, and Doordash. The "add to order" feature allows customers to append items to their existing orders without incurring additional shipping fees before a specified cutoff time. This study investigates the balance between immediate revenue increases and potential long-term profitability impacts. The central questions addressed include the overall value of the "add to order" feature to the platform and the optimal adjustment of free shipping thresholds and shipping fees following its implementation.

4 - An Exact Solution for Chance-Constrained Two Echelon Vehicle Routing Problem with Autonomous Delivery Robots

Dan Liu, Kean University, Union, NJ, United States, EVANGELOS KAISAR

This paper investigates the challenges of planning and operating an efficient E-grocery delivery system utilizing Autonomous Delivery Robots (ADRs) during peak demand periods. ADRs, critical in facilitating last-mile deliveries from vans at satellite nodes, operate within a structured two-echelon network. This study addresses key operational constraints such as customer delivery deadlines, and ADR energy limitations. We formulate a two-echelon vehicle routing problem that incorporates stochastic service times and demands. To tackle this problem, we employ a chance-constrained model and develop an efficient solution approach based on column generation. Central to our approach is a novel labeling algorithm that integrates two strategies for constructing the routes. To enhance the solution's reliability, we apply statistical inference techniques to ensure compliance with the chance constraints. Our results demonstrate the effectiveness of the stochastic formulation in reducing costs, showcasing significant improvements over traditional methods.

5 - Temporary medical shelter location-routing problem with manpower allocation in disaster situation

Joonghoo Park, Korea University, Seoul, Korea, Republic of, Hyungjoo Cha, San Kwon, Taesu Cheong

In this study, we focus on developing an efficient algorithm for medical operations system. We present a mathematical programming model addressing temporary medical shelter location-routing problem with manpower allocation (TMSLRPwMA). The proposed model integrates the ambulance routing problem and the temporary medical relief shelter location problem by considering the healthcare staffing requirements of patients. It is novel in the sense that we consider number of rescue staffs required in response to the patient demand. We aim to minimize the response time under patient's status, together with the staff allocations and vehicle usages. The concomitant problem is NP-hard, as TMSLRPwMA is a general case of the location-routing problem, and thus we design a hybrid genetic algorithm to retrieve good-enough solutions under a reasonable calculation time. Computational results under different scenarios are provided to demonstrate the validity of the proposed algorithm for various situations.

6 - Locomotive Routing for Torpedo Ladle Car Transportation in the Steel Industry

Ohhyun Kweon, POSTECH, Pohang, Korea, Republic of, Byung-In Kim

The locomotive routing problem (LRP) deals with the optimization of locomotive routes for transporting torpedo ladle cars in the steel industry. LRP belongs to a family of pickup and delivery problem time windows with Last-In First-Out constraints. We extended LRP considering many-to-one and one-to-many requests, along with additional facilities and practical constraints. Two mathematical models and a matheuristic algorithm based on logic-based Benders decomposition approach are proposed. For the 20 randomly generated small-sized instances, the mathematical model could mostly provide solutions optimally, while the matheuristic algorithm provided near-optimal solutions. In the case of 20 medium- and large-sized instances, the mathematical model failed to find a solution, whereas the matheuristic algorithm generated good feasible solutions, outperforming Adaptive Large Neighborhood Search algorithm for all instances.

7 - Heuristic Strategies for Optimizing MDVRP Bridging Transshipment and Assignment Problems in Supply Chain Logistics

Pitchaya Pongmanawut, Chiang Mai University, Chiang Mai, Thailand, Chulin Likasiri

In this work, we address the Multiple Depot Vehicle Routing Problem (MDVRP), a complex optimization task involving depots, suppliers, and customers. Our problem includes a general structure in which routes start at depots, pass through suppliers to gather raw materials, then return to depots to be transformed before delivering the final product to customers. We decompose the problem into transshipment and assignment sub-problems, tackling them sequentially to optimize resource utilization and minimize operational costs. We propose heuristic methods designed for the MDVRP setting in order to tackle the natural complexity. The algorithm is made to minimize overall distance traveled and match vehicle constraints, all the while navigating the complex network of suppliers, depots, and customers in an efficient way. Our approach is continually compared with current methods after being established through considerable experimentation on benchmark instances. The results demonstrate the effectiveness of our proposed heuristics in achieving competitive solutions for real-world supply chain logistics optimization.

Keywords: MDVRP, transshipment problem, assignment problem, heuristics

8 - Heuristic Approaches for Vehicle Routing Problems with Collaborative Multiple Heterogenous Robotic Systems in Complex Scenarios

Hyeseon Lee, Michigan Technological University, Houghton, MI, United States, Abhishek Patil, Myoungkuk Park, Jung Yun Bae

Finding efficient task allocation and routes, known as the vehicle routing problem (VRP), is crucial for the operations of multi-robot systems, particularly in scenarios involving heterogeneous agents with collaborative motion. When applications involve large and complicated scenarios, prioritizing a solution that offers a good approximation within a reasonable timeframe proves more beneficial than striving for an exact solution. This research addresses this necessity by introducing two heuristic approaches capable of resolving VRPs with various constraints.

The first approach utilizes the Large Language Model (LLM) to tackle generalized problems. We initially provide feasible solutions to small-size problems to the LLM, optimizing the prompts with detailed descriptions of task attributes, constraints, and environmental conditions. By guiding the LLM through iterative prompt refinement, it learns the user requirements and generates feasible routes for each robot while minimizing the objective function. On the other hand, the second heuristic approach adopts a greedy strategy, iteratively enhancing solution quality from an initial feasible solution while satisfying all pertinent constraints. While the LLM-based approach accommodates a broader spectrum of problems, its tendency to produce infeasible solutions necessitates caution. In contrast, the heuristic approach offers stability, guaranteeing feasibility by incorporating all constraints within its mathematical formulations.

Both approaches are implemented and tested in simulations with multiple scenarios, varying the number of tasks, robots, and constraints to validate the effectiveness of the proposed methods. The computational results are presented, showing the potential of the approaches to be utilized in real-world applications.

SE13

Summit - 333

Financial and Risk Analysis

Flash Session

Flash

Chair: Devin Zhang, 352 Business Building, 16802

1 - The Best Time to Play the Lottery

Christopher Rump, Bowling Green State University, Bowling Green, OH, United States

The best time to play the lottery is when the jackpot has rolled over several times and grown large, but not so large that you must share the prize if you win. We examine maximizing the expected value of a winning ticket as well as that in a random ticket. The derived optimality criteria depend on the prize elasticity of ticket demand. A regression analysis on data obtained from the Mega Millions® and Powerball® multistate lotteries suggests ticket sales grow quadratically in the size of the advertised lump-sum cash jackpot prize. With quadratic growth, the best time to play is when ticket sales are 1.25 to 2.5 times the jackpot odds, currently about 300 M to one for these two lotteries. Since ticket sales are not known to ticket buyers, we invert the regression function to prescribe the best time to play in terms of the cash prize. It turns out that these lotteries offer a (pre-tax) fair wager with positive expected value in a surprisingly wide interval of jackpot prizes. That is a good time to play; the best time is in the neighborhood of the nearly 1 $B record cash jackpot awarded in these lotteries in recent years.

2 - Credit risk guarantee and contagion in supply chains

Hariprasad Bellamkonda, Indian Institute of Management-Indore, Indore, India

The diverging interests of suppliers and buyers about trade-credit maturity are to be resolved using supply chain finance programs. While these programs relax suppliers' liquidity constraints for small- and medium-sized enterprises, the credit risk shifted to credit guarantee networks thereby the potential risk contagion increases. A payment-and-clearing model based on capital-constrained firms and suppliers has been examined. When certain event risks occur, the proposed debt-payment and asset-clearing strategy improves their payment ability and minimizes suppliers' clearing losses. While the credit guarantee networks present in the supply chain system can facilitate complex guarantee relationships, the guarantee financing scheme under the linear threshold model can minimize risk spillover. The study enhances understanding of effectiveness and efficiency of the schemes in diminishing systemic risks thereby providing sustainable credit arrangements in the supply chains.

Keywords: risk analysis, Credit guarantees, Risk contagion, Asset clearing

3 - You Are What You Like： the Relations Between Daily Preferences and Higher Order Risk Attitudes

Peng Du, Southwestern University Of Finance And Economics, Chengdu, China, People's Republic of, Lin Zhao, Hua Chen

Higher-order risk preferences, including risk aversion, prudence, and temperance, play a fundamental role in economic decision-making, including daily consumption/saving behaviors, portfolio allocation, and retirement planning. Their importance stimulates a growing interest to gauge individuals' risk preferences through easily observable and reliable data. Extant literature has suggested that some daily preferences like diets and sports can reflect an individual's risk preferences, but such relations have not been thoroughly investigated and existing evidence is restricted to the second-order risk preference. We, therefore, conduct a systematical survey including thirty daily preferences across six categories (food, clothing, exercise, reading, music, and video), and three lottery-choice experiments to measure subjects' higher-order risk preferences. Our results indicate that, after controlling for individual characteristics, some daily preferences reveal strong associations with higher-order risk preferences. We also find that cognitive ability and personal traits are important mediating factors. We are the first to systematically study the relations between daily preferences and higher-order risk preferences. Additionally, our research provides a valuable demonstration of identifying risk preferences through mining individuals' behavioral data in the big data era.

4 - Corporate default prediction using advanced machine learning techniques

Stefanos Delikouras, University of Miami, Miami Herbert Business School, Coral Gables, FL, United States, Erotokritos Skordilis, Emily Struble

Corporate default, i.e., the inability of a corporation to meet its debt obligations, is one of the most important corporate events both for corporate finance and asset pricing. From a corporate finance perspective, the likelihood of default determines the debt capacity of a firm and the importance of debt in its capital structure (debt vs. equity). From an asset pricing standpoint, the probability of default dictates the fair value of debt securities. Ultimately, the likelihood of default affects the firm's cost of capital, i.e., the ease with which corporations can engage in new investments and labor hiring. Despite its paramount importance, the literature has not yet provided an accurate (both in- and out-of-sample) methodology to predict default. This is mainly due to the fact that corporate defaults, especially for highly-rated firms that constitute the majority of debt securities happen very scarcely: < 0.2% for investment rated bonds, 1.1% for BB-rated, 4.8% for B-rated bonds. In this work, we utilize state-of-the-art machine learning techniques on a large database of >22,000 companies to accurately predict corporate default from accounting information despite the sparsity of default events.

5 - Catastrophe Risk Sharing among Individuals, Private Insurance, and Government

Hanyang Wang, Indiana University Kelley Business School, Bloomington, IN, United States, Ruo Jia, Jieyu Lin, Michael Powers

Limited research has been conducted on optimal public-private risk-sharing programs for catastrophe insurance markets. In the present work, we employ a game-theoretical framework to study the risk-sharing interactions of three types of catastrophe-market participants: a continuum of individuals, a private primary insurer, and a government that can choose between alternatives of (re)insurance or ex-post relief. Our analysis shows that the optimal government-intervention program varies depending on the correlation structure among individual catastrophe-loss exposures. For moderately positive levels of correlation, it is optimal for government to offer an ex-post relief program to supplement private insurance, but for higher levels of correlation, government reinsurance becomes optimal (although not to the extent of replacing the private market if government is less efficient than private firms). In sum, as catastrophe-loss correlations increase, more risk-sharing channels and funding are needed to maximize social welfare.

6 - CRADLE: Cyber Risk Assessment and Disclosure Language Evaluator

HILAL PATACI, University of Texas at San Antonio, San Antonio, TX, United States, Shen Chang, T Ravichandran

Effective corporate disclosure is crucial for informed stakeholder engagement. However, the pervasive use of standardized, boilerplate language, a form of discretionary behavior often hides critical information, increasing risks for investors and the broader economy. This issue is particularly acute in cyber risk reporting within annual statements, which are typically repetitive and vague. Despite calls from the Securities and Exchange Commission (SEC) for more transparent cybersecurity disclosures and the upcoming requirement for a dedicated Cyber risk factors section by 2024, the entrenched practice of using opaque language continues, often driven by economic motives. Employing principles from deterrence theory and integrating the National Institute of Standards and Technology's (NIST) Risk Assessment Framework, we developed CRADLE: Cyber Risk Assessment and Disclosure Language Evaluator, a novel method to detect and eliminate generic cybersecurity risk disclosures. In this work, we ask the question of how economic incentives and competitive pressures by shaping the persistence of boilerplate language impact the litigation outcomes for companies. We argue that merely introducing a (1.C) cybersecurity risk section in annual statements may not be sufficient due to the strategic use of boilerplate language for economic incentives and competitive pressures Further, this improved understanding of the consequences associated with discretionary behavior in cybersecurity risk disclosures aims to inform regulatory policies and corporate governance practices, to foster more transparent and effective risk disclosure.

7 - A Model for Determining Threshold Guarantee Levels in Supported Infrastructure Projects

Luiz Brandao, PUC-Rio, Rio de Janeiro, Brazil, Glaucia Fernandes, Naielly Marques, James Dyer

There is a growing reliance on private capital for funding infrastructure projects globally, particularly through Public-Private Partnerships (PPPs). However, private investors may perceive certain projects as excessively risky, potentially leaving governments without viable bidders. To incentivize private investment, governments often provide risk mitigating mechanisms such as subsidies, incentives, revenue guarantees or even term extensions. The Minimum Revenue Guarantee (MRG) is a common risk mitigating mechanism in PPPs, but the optimal procedure to determine its optimal design remains unclear. Building on existing literature, this paper proposes the use of the cap and floor model, widely used in European electricity markets, to mitigate risks in infrastructure projects in transportation. Unlike revenue-based approaches, the model's cost-based cutoff values offer simplicity and transparency. The floor guarantees minimum revenues to cover project costs, while the cap restricts excessive returns. Although successfully applied in electricity markets, its application to transportation infrastructure remains underexplored. The paper contributes to the literature by showing how this approach can be successfully applied to infrastructure financing. We model the cap and floor regime under the real options approach and apply this model to a numerical example.

Our results suggest that this model is effective in assisting policy makers determine the optimal cap and floor levels while minimizing the cost to the government and taxpayers

8 - Developing a potential assessment framework for startups: Bridging the gap between traditional methods and venture capital needs

woojin cho, seoul national university, seoul, Korea, Republic of, sungjoo lee

Corporate potential assessment is a critical area in economics and management, particularly important for startups in their early stages. Accurate assessment of a startup's potential is crucial for securing venture capital and devising long-term growth strategies. However, startups pose unique challenges for potential assessment due to their innovative nature, generally low revenue, and frequent operational losses. Traditional methods often prove inadequate for assessing the potential of these startups because of their distinct financial profiles.

This study seeks to address these challenges by developing a potential assessment framework specifically designed for startups. It will explore the key factors that venture capitalists consider when evaluating early-stage companies, including market potential, technological innovation, and the competitive environment. This analysis aims to provide a nuanced approach that caters to the unique needs and circumstances of startups. It will facilitate more informed investment decisions by venture capitalists, enhancing the efficacy of their evaluations.

By proposing a new approach to assessing startup potential, the anticipated impact of this research is significant. It aims to provide a clearer, more adaptable assessment framework that could contribute substantially to the field of venture capital and entrepreneurship. The proposed framework is expected to encourage more strategic and judicious investment practices, bridging the gap between traditional assessment methods and the dynamic, often unpredictable startup ecosystem. This could lead to more sustainable business ventures and a healthier investment climate, ultimately benefiting the broader economic landscape.

9 - The Working Capital Channel of Investment under Uncertainty

Devin Zhang, Pennsylvania State University, State College, PA, United States, Devin Zhang

This study documents a novel working capital channel of investment under uncertainty, wherein uncertainty affects investment not through adjustment costs or irreversibility but through working capital and cash flows. The uncertainty comes from a prevalent supply chain phenomenon known as the bullwhip effect (the BWE), where demand volatility is amplified upstream along the supply chain from retailers to raw materials suppliers. Analyzing a sample of over 155,000 supply chains, I find that a high bullwhip effect leads firms to increase their inventory levels, raising working capital demands and reducing cash available for investment, especially for financially constrained firms. This channel links the investment under uncertainty and investment-cash flow sensitivity literatures by showing how uncertainty can be transmitted through working capital and cash flow, ultimately impacting investment decisions. Furthermore, while extensively studied in supply chain management, the impact of the bullwhip effect on corporate finance remains largely unexplored. My study is also one of the first to explore how this well-known supply chain issue affects corporate finance and firms' investment decisions.

SE14

Summit - 334

INFORMS Finance Student Paper Competition

Award Session

Finance

Chair: Agostino Capponi, Columbia University, New York, NY, United States

SE15

Summit - 335

Resource Allocation Beyond Standard Assumptions

Invited Session

Revenue Management and Pricing

Chair: Chara Podimata, MIT, Boston, MA, United States

Co-Chair: Zijie Zhou, MIT, 70 Pacific Street, Apt 817, Cambridge, MA, 02139, United States

1 - Allocating Divisible Resources on Arms with Unknown and Random Rewards

Ningyuan Chen, Rotman School of Management, Toronto, ON, Canada, Wenhao Li

We consider a decision maker allocating one unit of renewable and divisible resource in each period on a number of arms. The arms have unknown and random rewards whose means are proportional to the allocated resource and whose variances are proportional to an order b of the allocated resource. In particular, if the decision maker allocates resource A_i to arm i in a period, then the reward is $Y_i=A_i\mu_i+A_i^b \xi_i$, where μ_i is the unknown mean and the noise ξ_i is independent and sub-Gaussian. When the order b ranges from 0 to 1, the framework smoothly bridges the standard stochastic multi-armed bandit and online learning with full feedback. We design two algorithms that attain the optimal gap-dependent and gap-independent regret bounds for $b\in[0,1]$, and demonstrate a phase transition at $b=1/2$. The theoretical results hinge on a novel concentration inequality we have developed that bounds a linear combination of sub-Gaussian random variables whose weights are fractional, adapted to the filtration, and monotonic.

2 - Multi-Armed Bandits with Network Interference

Justin Whitehouse, Carnegie Mellon University, Pittsburgh, PA, United States, Abhineet Agarwal, Anish Agarwal, Lorenzo Masoero

We consider a multi-armed bandit (MAB) problem where an agent sequentially assigns one of A possible actions to N units across T rounds. Unlike traditional MAB problems, the reward of each unit depends on the treatments assigned to other units, i.e., there is interference across the underlying network of units. Online experimentation with network interference is a common challenge in applications such as e-commerce and medicine. Even though there is a large literature on performing statistical inference with interference in the offline setting, far less is known about how to adaptively assign treatments to maximize average reward. With A actions, and N units, minimizing regret is combinatorially difficult since there is a total of A^N actions for the agent to explore. To overcome this issue, we impose a natural and widely-assumed structure on the interference pattern, where the reward of a unit is only affected by the treatments assigned to s neighboring units. Under this model, we use tools from discrete Fourier analysis to develop a sparse linear representation of the underlying rewards, and propose a simple Lasso-based algorithm to estimate this function and minimize regret. Our algorithms can not only take advantage of the interference structure if it is known, but learn the underlying network structure when it is unknown. Our paper significantly generalizes other works on MABs with network interference, which impose stricter conditions on a known interference pattern, and also compare regret to a markedly weaker optimal action.

3 - Regulating Discriminatory Pricing in the Presence of Tacit Collusion

Zongsen Yang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Xiao Lei, Pin Gao

Price-setting algorithms have facilitated widespread awareness-based price discrimination, in which firms charge high prices to customers unaware of alternative choices and low prices to those in competitive markets. This unethical behavior has increased customer complaints and prompted policymakers to enact fairness regulations in response. However, while limiting price discrimination may improve consumer welfare under genuine competition, it also affects firms' incentives to form tacit collusion, another regulatory concern arising from the proliferation of pricing algorithms.
We develop an analytical model to examine the interplay between fairness regulation and tacit collusion, and discuss its impact on consumer welfare and policymaking. Firms utilize customers' product unawareness to implement price discrimination and decide whether to collude by comparing profits from collusion and deviation. We then explore the consequences of price fairness regulation on the sustainability of tacit collusion. For homogeneous products, fairness regulation can substantially weaken collusion, potentially rendering it unattainable. However, for differentiated products, strict fairness inadvertently supports collusive behavior, harming consumer welfare. In this case, mild fairness permitting moderate price differentiation can prevent market collusion and optimize welfare. We conduct a numerical experiment using simple Q-learning algorithms to demonstrate the realism of our analytical findings. To address fairness-induced collusion, we propose a novel approach that randomizes over the fairness levels, achieving the desired fairness in expectation. Our randomized policy is effective in both obstructing collusion and improving consumer surplus.
Overall, our study emphasizes the importance of a nuanced approach to regulating discriminatory pricing in the presence of tacit collusion.

4 - When Should you Offer an Upgrade: Online Upgrading Mechanisms for Resource Allocation

Zijie Zhou, MIT, Cambridge, MA, United States, Patrick Jaillet, Chara Podimata, Andrew Vakhutinsky

In this work, we study an upgrading scheme for online resource allocation problems. We work in a sequential setting, where at each round a request for a resource arrives and the decision-maker has to decide whether to accept it (and thus, offer the resource) or reject it. The resources are ordered in terms of their value. If the decision-maker decides to accept the request, they can offer an **upgrade-for-a-fee** to the next more valuable resource. This fee is dynamically decided based on the currently available resources. After the upgrade-for-a-fee option is presented to the requester, they can either accept it, get upgraded, and pay the additional fee, or reject it and maintain their originally allocated resource.

We take the perspective of the decision-maker and wish to design upgrading mechanisms in a way that simultaneously maximizes revenue and minimizes underutilization of resources. Both of these desiderata are encapsulated in a notion of **regret** that we define, and according to which we measure our algorithms' performance. We present a fast algorithm that achieves **O(log T)** regret. Finally, we implemented our algorithm utilizing data akin to those observed in the hospitality industry and estimated our upgrading mechanism would increase the annual revenue by over **17%**.

SE16

Summit - 336

Advances in Revenue Management: From Theory to Practice

Invited Session

Revenue Management and Pricing

Chair: Anran Li, The Chinese University of Hong Kong, Hong Kong

Co-Chair: Zhe Liu, Imperial College London, London, United Kingdom

1 - Managing Newsvendors via Demand Nudging

Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, China, People's Republic of, Yilun Chen

The increasingly prevalent "fulfillment by platform" practice in the digital economy presents platforms with the unique challenge of managing their third-party sellers' autonomous inventory decision-making. In this work, we explore a novel idea to mitigate this challenge through nudging demand among sellers, which reshapes their perceived demand and incentivizes them to make platform-favorable inventory decisions. To this end, we study a stylized Stackelberg game between a platform and a continuum of third-party sellers selling substitutable products on the platform. The platform aims to design a demand nudging policy to maximize her profit, while the sellers, in response, determine their inventory according to the post-nudge demand in a newsvendor manner. We establish in closed form the platform's optimal policy. It employs a simple cut-off structure called defluctuating and resembles risk pooling in the centralization setting where the platform can make joint demand-substitution and inventory decisions. Despite the power of this control on the off-equilibrium path, we show that it is Pareto-improving for both the platform and the sellers under a mild condition. Furthermore, when the products are substitutable enough, the platform can achieve the centralization benchmark. On the contrary, if the products are not substitutable enough, the platform's service level to the customers may, in certain primitive regimes, decrease.

2 - Assortment Optimization with Market Share Threshold under Multinimal Logit Choice Model

Wenchang Zhu, Cornell Tech, New York, NY, United States, Paat Rusmevichientong, Huseyin Topaloglu

We consider assortment optimization problem with market share threshold constraints under multinomial logit choice model.

3 - Dynamic Pricing Under Self-Exciting Arrival Processes

Longyuan Du, University of San Francisco, San Francisco, CA, United States, Ming Hu, Quan Yuan

We study dynamic pricing problems with a non-linear stochastic Hawkes arrival process where demand is influenced by previous customers. We prove that the seller would first set a low price to attract customers and increase the price over time. The optimal price increases in the exciting factor when the exciting function is concave and decreases in the exciting factor when the exciting function is convex. We further study easy-to-compute heuristics and their performances in the asymptotic regime. We show that the heuristic adopting the optimal price in the deterministic problem incurs a constant performance loss. The paper is extended to consider general information spreading and multiple identical products.

4 - Speed Service with Spotlight Products: Assortment Optimization and Pricing in a Queue

Zhe Liu, Imperial College London, London, United Kingdom, Kalyan Talluri, Shixin Wang

Quick-service restaurants, such as fast-food franchises and coffeehouse chains, recommend items with spotlight menu that may shorten customers' decision time and hence speed up the service, but with a trade-off of possibly lower profit margin. We study such assortment optimization problems when service rate is affected by the recommendation and plays a role in the revenue rate objective. Formulated as a sequential recommendation problem in a queueing setting under the multinomial logit model, we obtain structural properties of the impact of arrival rate and system capacity on optimal assortment, and find optimality conditions for attractiveness-ordered or fare-ordered nested policies. We also characterize the optimal joint assortment and pricing under common price sensitivity.

SE17

Summit - 337

How Competition Can Disrupt Machine Learning

Invited Session

Revenue Management and Pricing

Chair: Meena Jagadeesan, UC Berkeley, Berkeley, CA, United States

Co-Chair: Nika Haghtalab

1 - Impact of Decentralized Learning on Player Utilities in Stackelberg Games

Kathleen Donahue, Cornell, Ithaca, NY, United States, Nicole Immorlica, Meena Jagadeesan, Brendan Lucier, Alex Slivkins

When deployed in the world, a learning agent such as a recommender system or a chatbot often repeatedly interacts with another learning agent (such as a user) over time. In many such two-agent systems, each agent learns separately and the rewards of the two agents are not perfectly aligned. To better understand such cases, we examine the learning dynamics of the two-agent system and the implications for each agent's objective. We model these systems as Stackelberg games with decentralized learning and show that standard regret benchmarks (such as Stackelberg equilibrium payoffs) result in worst-case linear regret for at least one player. To better capture these systems, we construct a relaxed regret benchmark that is tolerant to small learning errors by agents. We show that standard learning algorithms fail to provide sublinear regret, and we develop algorithms to achieve near-optimal O(T^{2/3}) regret for both players with respect to these benchmarks. We further design relaxed environments under which faster learning (O(T^{0.5})) is possible. Altogether, our results take a step towards assessing how two-agent interactions in sequential and decentralized learning environments affect the utility of both agents.

2 - How and Why to Manipulate Your Own Agent: On the Incentives of Users of Learning Agents

Yoav Kolumbus, Cornell University, Ithaca, NY, United States, Noam Nisan

The usage of automated learning agents is becoming increasingly prevalent in many online economic applications such as online auctions and automated trading. Motivated by such applications, we study the fundamental modeling and analysis of the strategic situations that the users of automated learning agents are facing. We consider strategic settings where several users engage in a repeated online interaction, assisted by regret-minimizing learning agents that repeatedly play a "game" on their behalf. We propose to view the outcomes of the agents' dynamics as inducing a "meta-game" between the users. Our main focus is on whether users can benefit in this meta-game from "manipulating" their own agents by misreporting their parameters to them. We define a general framework to model and analyze these strategic interactions between users of learning agents for general games and analyze the equilibria induced between the users in several classes of games, including repeated auctions. We show that, in many games, users have incentives to misreport their parameters to their own agents, and that such strategic user behavior can lead to very different outcomes than those anticipated by standard analysis.

The talk is based on two joint works with Noam Nisan:

https://arxiv.org/pdf/2112.07640 (NeurIPS 2022)

https://arxiv.org/pdf/2110.11855 (WWW 2022)

3 - Algorithmic Collusion by Large Language Models

Sara Fish, Harvard University, Cambridge, MA, United States, Yannai Gonczarowski, Ran Shorrer

The rise of algorithmic pricing raises concerns of algorithmic collusion. We conduct experiments with algorithmic pricing agents based on Large Language Models (LLMs), and specifically GPT-4. We find that (1) LLM-based agents are adept at pricing tasks, (2) LLM-based pricing agents autonomously collude in oligopoly settings to the detriment of consumers, and (3) variation in seemingly innocuous phrases in LLM instructions ("prompts") may increase collusion. These results extend to auction settings. Our findings underscore the need for antitrust regulation regarding algorithmic pricing, and uncover regulatory challenges unique to LLM-based pricing agents.

4 - Understanding Competition-Driven Content Ecosystems

Fan Yao, University of Virginia, State College, PA, United States

The ever expanding of content creation platforms has nurtured a booming of digital content ecosystems, which are driven by a competitive dynamics among content creators vying for user attention or revenue. This competitive environment poses significant challenges to the health of content ecosystems, including the risk of diminishing content diversity and the potential pursuit of popularity may drive out high-quality content. Our research establishes a game-theoretical framework to analyze these dynamics by viewing creators as strategic agents aiming to maximize platform-allocated resources. We will assess how platforms' algorithmic decisions and revenue allocation models impact content distribution and social welfare, providing insights for platform designers to enhance content quality and fairness, thereby fostering a healthier and more sustainable digital content ecosystem. We will also discuss how the advent of Generative AI might influence the content ecosystem and possible interventions to mitigate the potential adverse effects.

5 - Algorithmic Content Selection and the Impact of User Disengagement

Eric Zhao, University of California, Berkeley, Berkeley, CA, United States, Emilio Calvano, Nika Haghtalab, Ellen Vitercik

The content selection problem of digital services is often mathematically modeled as a decision-process where a service chooses, over multiple rounds, an arm to pull from a set of arms that each return a certain reward. This classical model does not account for the possibility that users disengage when dissatisfied and thus fails to capture an important trade-off between choosing content that promotes future engagement versus immediate reward. In this talk, I will discuss a new model for the content selection problem where dissatisfied users may disengage and where content that maximizes immediate reward does not necessarily maximize the odds of future user engagement. I will share theoretical results on the learnability of optimal content selection policies in this model, including efficient planning algorithms and online learning algorithms for serving users with unknown engagement patterns. I will also discuss the use of this model for identifying key primitives that determine how digital services should weigh engagement against revenue. For example, when it is more difficult for users to rejoin a service they are disengaged from, digital services naturally see a reduced payoff but user engagement may—counterintuitively—increase.

SE18

Summit - 338

Learning and Optimization in Risk and Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Zhenyu Hu, National University of Singapore, Singapore, Singapore

Co-Chair: Yifan Feng, National University of Singapore, Singapore, Singapore

1 - Customer Lifetime Value and Algorithmic Pricing

Anton Ovchinnikov, Smith School of Business, Queen's University, Kingston, ON, Canada, Jue Wang

Customer Lifetime Value (CLV) is commonly defined as “the present value of the future cash flows attributed to the customer relationship,” Pfeifer et al. (2005). A large body of CLV literature focused on retention, loyalty, and other nuances of how to calculate “the present value of the future,” but how much value is obtained from each customer transaction received little attention: most studies adopted the aforementioned cash-flow-based approach. In this paper we show that certain pricing algorithms utilized across many industries such as transportation, hospitality, retail, advertising, and sports & entertainment, make the cash-flow-based approach systematically biased, overvaluing high-price transactions and undervaluing low-price ones. We prove that the bias depends on the elasticity properties of the demand function used by the firm’s pricing algorithms. Building on that, we also show that certain non-stationary customer arrival patterns could restore the “intuitive” directional relationship that higher-priced transactions are more valuable. Lastly, we present an intuitive refined approach to corrects for this bias.

2 - Credit Rating Design Under Adverse Selection

QINZHEN LI, National University of Singapore, Singapore, Singapore, Yifan Feng, Jussi Keppo

We introduce a model of the credit rating market where investors depend on Credit Rating Agency's (CRA) ratings to predict firms' default probabilities. Firms seeking funding may pay the CRA for solicited ratings. The revenue-maximizing CRA strategically designs a rating system based on each firm’s quality and solicitation action. We find that in equilibrium, high-quality firms tend to solicit ratings more aggressively and receive higher ratings than low-quality ones. Although solicited ratings are generally higher than unsolicited ones, this difference diminishes when controlling for firms' true underlying qualities. Our model suggests that the reported discrepancy between unsolicited and solicited ratings in the existing literature can be primarily due to adverse selection on the demand side of the credit rating market. Additionally, we conduct robustness checks to validate variations of our model.

3 - Fragility-Aware Classification for Understanding Risk and Improve Generalization

Chen Yang, Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Jin Qi, Ruohan Zhan

In various applications, including medical diagnosis and default detection, the classification model is applied, and managers make decisions based on the classifier's predictions. Many performance metrics have been developed by comparing the actual label and the classifier's prediction. However, the confidence of the model prediction is also essential and neglected by commonly used performance metrics. Confidence represents the risk and uncertainty of the model prediction, and managers may consider conducting different operations in instances with different confidence. Due to the growing complexity of learning models, state-of-the-art classification models like neural networks are known to be more accurate but suffer from the overconfidence issue, which can lead to a high risk in decision-making as the models cannot indicate when they are probably incorrect. These two points invoke the need to understand and regulate the model's confidence. In this paper, we propose a novel metric, the Fragility Index (FI), to evaluate the risk of wrong prediction by capturing the magnitude of the ranking error from a risk-averse perspective. To strengthen the generalization ability, we adopt the robust satisficing (RS) framework to design FI. We show that FI can identify the information that conventional metrics like accuracy cannot capture. On top of that, we design a model training framework based on FI and RS. We demonstrate the effectiveness of FI and FI-based classification models using synthetic data, real data, and deep neural networks. With FI, the decision makers can understand the risk inside the learning model predictions and make better decisions.

4 - Multi-Objective Assortment Optimization: Profit, Risk, Customer Utility, and Beyond

Zhen Chen, Arizona State University, Tempe, AZ, United States, Heng Zhang, Hongmin Li, Scott Webster

Assortment optimization is a fundamental problem in revenue management, where the goal is usually to select a subset of products to offer to customers in order to maximize expected revenue or profit. Indeed, although revenue is of predominant importance, it is rarely the sole objective of a firm's product assortment decisions. Business practices often involve multiple, and potentially conflicting goals. In this work, we propose a general framework and a novel reformulation method for solving multi-objective assortment optimization problems. We focus on problems with a separable sum of multiple convex objective functions on linear combinations of choice probabilities, presenting a reformulation that effectively "linearizes" the problem. We prove that the reformulated problem is equivalent to the original and leads to a unified solution approach for multi-objective assortment optimization problems in various contexts. Our approach encompasses a wide range of operational objectives, such as risk, customer utility, market share, costs with economies of scale, and dualized convex constraints. We first illustrate our method using the multinomial logit model, both without constraints and with totally unimodular constraints. We further show that our framework provides to tractable solutions under the nested logit model and the markov chain choice model. Together with large-scale numerical experiments to demonstrate the efficiency and practicality of our methods, we highlight that our work provides a powerful and flexible tool for solving multi-objective assortment problems, which arise frequently in practical revenue management settings.

SE19

Summit - 339

Pricing and Yield Management at Amazon

Invited Session

Revenue Management and Pricing

Chair: Ozge Sahin, Johns Hopkins Carey Business School, Baltimore, MD, United States

Co-Chair: Samyukta Sethuraman, Amazon, Folsom, CA, United States

Co-Chair: Samyukta Sethuraman, Amazon, Folsom, CA, United States

1 - Early Stopping Methods at Pricing Lab

Sid Sanghi, Amazon.com, Seattle, WA, United States, Daria Zelenia, James Nordlund, Mohsen Bayati

In this paper we implement two methods to enable early lab terminate once enough evidence of strong positive or strong negative effect has been collected. We implement the approach on historic labs and find that both methods could help save 50% of experiment duration.

2 - Price Optimization with Neural Thompson Sampling with Dropout

Ru Wang, Amazon, Cupertino, CA, United States, Hyungjun Lee, Laleh Jalali

The single-product price optimization problem aims to find the optimal prices for Amazon while assuming that the demand of each product is independent of each other. We present a novel solution to the single-product price optimization problem using a neural network-based contextual bandit methodology called Neural Thompson Sampling with Dropout (NTSD). NTSD is a Thompson Sampling method that uses a neural network with dropout to estimate the reward distributions of the arms. It allows a neural network with any architecture to be used as the underlying reward estimation model, and provides computationally efficient methods for training the neural network estimator as well as sampling from the implied reward distributions of the arms. In simulation studies, NTSD generates 4-8% improvements in cumulative rewards over the baseline production model under simulations.

3 - Combining ML & OR for Efficient Amazon Locker Utilization

Samyukta Sethuraman, Amazon, Santa Clara, CA, United States

Amazon Locker is a self-service delivery or pickup location where customers can pick up packages and drop off returns. A basic first-come-first-served policy for accepting package delivery requests to lockers results in lockers becoming full with standard shipping speed (3- to 5-day shipping) packages, leaving no space for expedited packages, which are mostly next-day or two-day shipping. This paper proposes a solution to the problem of determining how much locker capacity to reserve for different ship-option packages. Yield management is a much-researched field with popular applications in the airline, car rental, and hotel industries. However, Amazon Locker poses a unique challenge in this field because the number of days a package will wait in a locker (package dwell time) is, in general, unknown. The proposed solution combines machine learning techniques to predict locker demand and package dwell time with linear programming to maximize throughput in lockers. The decision variables from this optimization provide optimal capacity reservation values for different ship options. This resulted in a year-over-year increase of 9% in Locker throughput worldwide during the holiday season of 2018, impacting millions of customers.

SE20

Summit - 340

Advanced Decision Technologies for Industry Applications

Invited Session

Decision Analysis Society

Chair: Kuo-Ping Lin, Tunghai University, Taichung City, Taiwan

1 - Evaluating Infant Development Through a Hybrid Intuitionistic Fuzzy Classification System

TING YU LIN, Tunghai University, Taichung City, Taiwan, YA CHI HSU, KUO PING LIN

This study proposes a hybrid approach using clustering algorithms and classification methods to enhance the accuracy and granularity of infant developmental assessments. The classification system adopts k-means determine the difference in infant development between 6 months and 1 year of age while the input datasets are percentage and over time, which are vague sets. The classification algorithm adopts long short-term memory network (LSTM) offer a robust framework for vagueness inherent in developmental assessments. Through a comprehensive evaluation using real-world infant development data, we demonstrate the superior performance of the proposed classification system. The hybrid classification system can achieve superior performance for cognitive scores which accurate rate is 0.95. This study contributes to ongoing efforts to advance assessment methodologies in early childhood development and underscores the potential of hybrid computational approaches in pediatric healthcare.

2 - Construct a novel SCAN method of unsupervised learning for dealing with the auto-annotation task in semiconductor manufacturing processes

CHIH-HUNG JEN, Lunghwa University of Science and Technology, Taoyuan City, Taiwan, SHU-KAI S. Fan, Yu-Ching Huang

Semiconductor manufacturing is developing rapidly, and related production technology has made significant progress. In practical semiconductor processes, defect analysis of wafer maps is a crucial step in enhancing product quality and yield. Additionally, these defect patterns can provide valuable process information, allowing process engineers to pinpoint the causes of process anomalies. However, in supervised learning, manually labeling data for wafer maps is a tedious and time-consuming task that can lead to misjudgments when performed over long periods. To address these shortcomings, this paper focuses on automatically grouping data into semantically meaningful clusters without real labels, a process known as pseudo-labeling. It utilizes a novel Semantic Clustering by Adopting Nearest-neighbors (SCAN) method to help the model learn data representations and extract meaningful features, reducing reliance on low-level features for subsequent clustering. Within the new SCAN model, a fine-tuning process generates pseudo-labels. The self-labeling model uses confident samples to correct the pseudo-labels of their nearest neighbors, gradually improving the accuracy of initially misclassified neighbors using the FixMatch loss function. Experimental results show that the auto-annotation system achieves significantly improved performance, potentially eliminating the need for extensive manual labeling.

3 - Applying Constraint Theory and Resource-Based Views: A Case Study of Problem-Solving in the Traditional Auto Parts Industry

Wu Sheng-pin, National Cheng Kung University/Department of Business Administration, Tainan, Taiwan, Lin Tai-Yu, Wang Chien-Chih

Since the onset of COVID-19, global business models have evolved significantly. This study explores these transformations within the context of supply chain management, focusing on transitioning from risk diversification strategies to centralized procurement. This strategic shift aims to empower suppliers by enhancing their negotiating power over raw materials and transportation costs. Our investigation centers on a multinational auto parts supplier, which, despite the predominance of price and purchase volume considerations, continues to face challenges in production scheduling and manpower deployment, primarily due to delays in raw material supply. This research identifies three critical areas of concern: procurement negotiation skills, personnel recruitment and the management capabilities of middle and senior managers. Employing the Theory of Constraints and a Resource-Based View, this analysis not only addresses the delays in production delivery but also reveals the underlying deficiencies in management capabilities at middle and senior levels. Consequently, this paper proposes targeted improvement plans to augment managerial effectiveness and streamline supply chain operations.

Keywords: supply chain management, theory of constraints, resource-based view

4 - Using Graph Convolutional Network to Solve the Uncapacitated Single Allocation p-Hub Median Problem

Kang-Ting Ma, National Dong Hwa University, Hualien, Taiwan, You-Qin Wang

Logistics networks need restructuring to handle increased transportation demands and distribution complexity with the rise of Omni-channel retail. Previous heuristic algorithms for solving the P-Hub Median Problem failed to obtain optimal solutions for complex networks with many transfer points. This paper proposes a graph convolutional network (GCN) method to solve the uncapacitated single allocation p-hub median problem. Unlike traditional CNNs limited to Euclidean data, GCNs can learn better solutions by propagating and aggregating node features. The study clusters nodes, identifies suitable transfer locations, tests large-scale random examples, and explores graph attention mechanisms and loss functions to improve model generalization and performance.

SE21

Summit - 341

Machine Learning and Data Analytics for Robust Decision-Making in Complex Systems

Invited Session

Decision Analysis Society

Chair: Faraz Dadgostari, Montana State University, Bozeman, MT, 59718, United States

1 - AI-Enhanced Early Warning and Environmental Decision Support Systems: Real-Time Management of Harmful Algal Blooms

Neda Nazemi, Montana State University , Bozeman, MT, United States

The increasing frequency, intensity, and complexity of climate-induced hazards necessitate robust and adaptive decision-making in risk management. Advances in Artificial Intelligence (AI), especially in machine learning and deep learning, offer scalable solutions for forecasting models. These models integrate multi-level, multi-frequency, and multi-source data, making them essential in addressing environmental challenges like harmful algal blooms (HABs), which threaten water quality, marine life, and public health globally. HABs are characterized by their sudden occurrences and complex triggers, presenting challenges to traditional forecasting methods that often fail to provide timely and accurate predictions. This research leverages deep learning to overcome these limitations, enhancing early warning systems and decision-support tools that facilitate proactive responses and help mitigate environmental and economic impacts. We have developed a deep-learning framework to forecast chlorophyll-a levels, a primary indicator of algal bloom formation. This framework supports short and mid-term decision-making, enabling real-time monitoring and management of bloom events. Its ability to provide actionable insights swiftly outperforms conventional methods, empowering stakeholders with enhanced responsiveness. Additionally, this paper proposes integrating these AI-driven tools into existing environmental management frameworks to boost resilience and sustainability. By improving decision-making processes, these tools enable adaptive strategies that effectively respond to dynamic environmental threats, fostering a more sustainable interaction between humans and their environment. This approach not only addresses the immediate challenge of managing HABs but also illustrates the broader applicability of AI in environmental risk management. It sets a pathway for future research in similar contexts, demonstrating significant potential for technological advancement in ecological conservation.

2 - Towards explainable artificial intelligence through expert-augmented supervised feature selection

Meysam Rabiee, University of Colorado Denver, Denver, CO, United States

This paper presents a comprehensive framework for expert-augmented supervised feature selection, addressing pre-processing, in-processing, and post-processing aspects of Explainable Artificial Intelligence (XAI).

3 - Graph Neural Networks as surrogates for simulation-based decision-making in infrastructure systems

Negin Alemazkoor, University of Virginia, Charlottesville, VA, United States

In recent years, there has been a growing emphasis on integrating uncertainty into decision-making processes for civil infrastructure systems. This involves calculating expected costs and failure probabilities through high-fidelity simulations, which can be computationally intensive due to high-dimensional uncertainty. To mitigate this, various studies have explored using analytical surrogates like polynomial chaos expansion, radial basis function, and Gaussian processes, which have shown promise in accelerating probabilistic analysis under specific conditions. However, there's a rising interest in utilizing machine-learning models as surrogates due to their ability to capture complex relationships within uncertain inputs and quality of interest (QoI). Nonetheless, this approach requires substantial data for training and meticulous calibration of hyperparameters and model architecture, which can be challenging when estimating numerous QoIs across a system. Additionally, calling a large number of surrogate models within the analysis can lead to error aggregation and complicate result interpretation. Graph neural networks (GNNs) offer a potential solution by providing node- and link-level approximations, reducing the need for multiple surrogate models. Despite this, their application in civil infrastructure systems remains limited, especially for large-scale systems under high-dimensional uncertainty. This work aims to address the challenges of training GNN surrogates for large-scale CIS in probabilistic analysis. Specifically, we propose a novel multi-fidelity approach that substantially reduces the computational cost of surrogate development. We demonstrate its advantages by applying it to the optimal power flow problem under high-dimensional uncertainty.

4 - Verification, Validation, Assurance, and Trust of Machine Learning Models and Data for Safety Critical Applications

Shane Hall, Montana State University, Bozeman, MT, United States

As the use of machine learning (ML) models proliferates in commercial and defense applications, military organizations face significant challenges in evaluating the effectiveness, robustness, and safety of these ML models in military systems. Relying on ML-informed recommendations and decisions in these systems requires very high confidence that any resultant behaviors will fall within intended operational and mission bounds. Ensuring reliable and safe behaviors involves both ensuring accurate and comprehensive data is used in the creation and training of these ML systems and that the ML models are robust, accurate, and appropriately behaviorally bounded when employed using real data in military operations. ML models come in many forms, and the technologies used to create them are rapidly evolving, and hence, military organizations need 1) a process and framework to assess and measure the quality of training data and identify shortcomings that may lead to poorly trained ML models, and 2) a process and tools for ML model exploration that can assure confidence of model behavior within defined data boundaries and can also identify unintended or poor behavior in ML models if they exist. This presentation outlines the existing literature on the metrics and measures used to verify and validate (V&V) ML training data and models and describes the process, framework, and tools to analyze these metrics and measures. Results that demonstrate these metrics, measures, framework, and tools are provided for an open-source classification ML model and an autonomous vehicle reinforcement learning (RL) model.

5 - A Multi-Agent Inverse Sequential Decision Learning for Causal Behavioral Learning

Faraz Dadgostari, Montana State University, Bozeman, MT, United States

Understanding the emergent patterns and dynamics within complex adaptive systems, such as social or economic networks, requires analytical tools capable of modeling intricate behavioral interactions and uncovering underlying causal mechanisms. In this paper, we propose a multi-agent inverse sequential decision learning framework that addresses the limitations of existing approaches. Our framework integrates advanced decision-making theories from behavioral economics, mathematical neuroscience, and computational psychology, extending traditional inverse reinforcement learning (IRL) methodologies with novel mathematical formulations and algorithmic innovations. The proposed framework is uniquely designed to efficiently model heterogeneous agent behaviors, capture their adaptive responses to dynamic environmental stimuli, and infer the underlying reward functions driving their decisions. We demonstrate its ability to analyze causal relationships within the system and provide interpretable insights that go beyond the capabilities of rationality-based or simulation-focused methods.

SE22

Summit - 342

Integrating Generative AI with Sequential Decision-Making: Theoretical Advances and Practical Applications (I)

Invited Session

Decision Analysis Society

Chair: Yingru Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Ming Yin, Princeton University, Princeton, NJ, United States

1 - Training Language Model Agents via Hierarchical Multi-Turn RL

Andrea Zanette, Carnegie Mellon University, Pittsburgh, PA, United States

A broad use case of large language models (LLMs) is in goal-directed decision-making tasks (or "agent" tasks), where an LLM needs to not just generate completions for a given prompt, but rather make intelligent decisions over a multi-turn interaction to accomplish a task (e.g., when interacting with the web, using tools, or providing customer support). Reinforcement learning (RL) provides a general paradigm to address such agent tasks, but current RL methods for LLMs largely focus on single-turn rewards. Most single-turn RL methods cannot endow LLMs with the ability to intelligently seek information over multiple turns, perform credit assignment, or reason about their past actions -- all of which are critical in agent tasks. This raises the question: how can we design efficient multi-turn RL algorithms for LLMs? We develop a framework for building multi-turn RL algorithms for fine-tuning LLMs, that preserves the flexibility of existing single-turn RL methods for LLMs (e.g., proximal policy optimization), while accommodating multiple turns, long horizons, and delayed rewards effectively. To do this, our framework adopts a hierarchical RL approach and runs two RL algorithms in parallel: a high-level off-policy value-based RL algorithm to aggregate reward over utterances, and a low-level RL algorithm that utilizes this high-level value function to train a token policy within each utterance or turn. Our hierarchical framework, Actor-Critic Framework with a Hierarchical Structure (ArCHer), can also give rise to other RL methods. Empirically, we find that ArCHer significantly improves efficiency and performance on agent tasks, attaining a sample efficiency of about 100.

2 - Transferred Decoding for LLM Alignment

Ming Yin, Princeton University, Princeton, NJ, United States

Aligning foundation models is essential for their safe and trustworthy deployment. However, traditional fine-tuning methods are computationally intensive and require updating billions of model parameters. A promising alternative, alignment via decoding, adjusts the response distribution directly without model updates to maximize a target reward r, thus providing a lightweight and adaptable framework for alignment. However, principled decoding methods rely on oracle access to an optimal Q-function ($Q*$), which is often unavailable in practice. Hence, prior SoTA methods either approximate this $Q*$ using $Q\pi\mathtt{sft}$(derived from the reference $\mathtt{SFT}$ model) or rely on short-term rewards, resulting in sub-optimal decoding performance. In this work, we propose Transfer $Q*$, which implicitly estimates the optimal value function for a target reward r through a baseline model $\rho\mathtt{BL}$ aligned with a baseline reward $\rho\mathtt{BL}$ (which can be different from the target reward r). Theoretical analyses of Transfer $Q*$ provide a rigorous characterization of its optimality, deriving an upper bound on the sub-optimality gap and identifying a hyperparameter to control the deviation from the pre-trained reference $\mathtt{SFT}$ model based on user needs. Our approach significantly reduces the sub-optimality gap observed in prior SoTA methods and demonstrates superior empirical performance across key metrics such as coherence, diversity, and quality in extensive tests on several synthetic and real datasets.

3 - GPT-HyperAgent: Adaptive Foundation Models for Online Decisions

Yingru Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Zhi-Quan Luo

Foundation models pretrained on historical data often struggle with uncertainty in new and rare situations, leading to errors that require feedback for correction and policy refinement. We introduces GPT-HyperAgent, a scalable reinforcement learning algorithm designed for continual learning and adaptation in uncertain complex environments. GPT-HyperAgent leverages GPT for expressive feature embeddings and integrates HyperAgent as an uncertainty-aware reward model for exploration in contextual bandit problems with natural language inputs.

Theoretical analysis confirms HyperAgent achieves fast incremental uncertainty estimation with logarithmic per-step computational complexity, matching the performance of exact Thompson sampling in linear contextual bandits and Randomized Least-Square Value Iteration in tabular RL environments.

Empirical results confirms GPT-HyperAgent's effectiveness in real-world safety-critical applications, such as automated content moderation on digital platforms via human-AI collaborations.

We will also explore open questions about reinforcement learning with foundation models.

4 - Cost-Aware Decoding for Safe Language Model Deployment

Hao Liang, CUHK-SZ, Shenzhen, China, People's Republic of

In the rapidly evolving field of artificial intelligence, deploying large language models effectively and safely remains a formidable challenge. This talk presents "Cost-Aware Decoding," an innovative strategy designed to simultaneously optimize for performance and safety. This method integrates cost-aware mechanisms into the decoding process of language models, enabling dynamic evaluation and adjustment of responses to minimize potential harm without compromising utility.

The cornerstone of this approach is a novel cost model, crafted from human-aligned feedback, which assesses the potential harm of text outputs. This model facilitates real-time adjustments to the decoding strategy, ensuring that generated responses not only adhere to safety thresholds but also align with user intentions.

During the session, we will explore the development and integration of the cost model within the decoding framework, address the challenges associated with balancing harmlessness against helpfulness, and discuss the practical implications of deploying safer language models across various applications. Empirical validations of our method demonstrate its efficacy in enhancing the safety of AI outputs, setting the stage for wider and more responsible deployment of AI technologies.

SE23

Summit - 343

Multi-Criteria Decision-Making Under Uncertainty

Invited Session

Multi Criteria Decision Making

Chair: Shervin Zakeri, 7 rt. de Drize,, 1206

Co-Chair: Milosz Kadzinski, Poznan University of Technology, Poznan, Poland

1 - Review and experimental comparison of ranking and choice procedures for constructing a univocal recommendation in a preference disaggregation setting

Milosz Kadzinski, Poznan University of Technology, Poznan, Poland, Michal Wojcik

We account for the preference disaggregation setting given multiple criteria ranking and choice problems. An assumed preference model is a set of additive value functions compatible with the Decision Maker's pairwise comparisons of reference alternatives. The incompleteness of such indirect preferences implies the multiplicity of feasible functions and the ambiguity in indicating the most preferred alternative or ordering alternatives from the best to the worst. We review approaches that construct a univocal recommendation under such scenarios. They represent four groups of methods: procedures selecting a representative value function, decision rules, scoring methods, and mathematical models for constructing a robust ranking. The use of all thirty-five approaches is illustrated on a simple decision problem. Then, they are compared in an extensive computational study in terms of their abilities to reconstruct the DMs' true preferences and robustness of delivered recommendations given the support they are given in the set of all compatible models. The results are quantified in terms of seven performance measures. Their analysis indicates that in the context of choice, it is beneficial to consider the rank acceptabilities for the best ranks. For ranking problems, the most advantageous outcomes are attained by procedures that emphasize the most frequent relations or positions in the feasible polyhedron. Apart from the average results, we discuss how the performance of all approaches changes for different parameterizations of the decision problem and preference model.

2 - The Problem of Sensitivity Analysis in MCDA / MCDM

Evangelos Triantaphyllou, Louisiana State University, Baton Rouge, LA, United States, Juri Yanase

Multi-Criteria Decision Analysis (MCDA), also known as Multi-Criteria Decision Aiding, or Multi-Criteria Decision Making (MCDM), is a family of approaches that analyze a finite set of alternatives which are described in terms of a finite set of evaluative criteria. Perhaps the most crucial role in the implementation of such decision-making approaches is the use of the evaluative criteria. The reason is that almost always the weights of the evaluative criteria are subjective and thus hard to elicit accurately. Hence, performing a sensitivity analysis on the role these weights play in the determination of the final solution is a task that should be considered in any decision-making setting that involves such approaches. We analyze the sensitivity analysis problem from three perspectives. The first perspective is the type of the decision problem (also known as the “problematic”) to be considered. For instance, is the focus on the ranking of the alternatives? Or on determining the best alternative? The second perspective is the impact ranking changes of the alternatives have. The third perspective is the likelihood the decision maker(s) may consider different weight values than the ones assumed initially. This sensitivity analysis problem is proposed to be treated in a universal manner that is applicable to a wide range of MCDA / MCDM approaches. Early findings suggest that existing sensitivity analysis approaches may yield dramatically misleading results. If the quest is the development of smart decision-making approaches, then the proposed sensitivity analysis approach must become an integral component of such smart approaches.

3 - Evaluating Blockchain-IoT Adoption in Supply Chains: A Fuzzy Entropy-Based MCDM Approach for Enhancing Environmental Sustainability

Atefeh Shoomal, University of Texas at Arlington, Arlington, TX, United States, Mohammad Jahanbakht

In the context of modern supply chains, the integration of blockchain and Internet of Things (IoT) technologies presents a significant opportunity to enhance environmental sustainability. Our study adopts a Multi-Criteria Decision Making (MCDM) approach, grounded in the Technology-Organization-Environment (TOE) frameworks, to critically assess the factors that influence the adoption of Blockchain-IoT in supply chain. We initiate our investigation by conducting a Delphi study to achieve a consensus among industry experts regarding the critical criteria for technology adoption, emphasizing environmental sustainability, technological readiness, and organization maturity.

Following the expert consensus, we employ a fuzzy entropy method to evaluate these criteria. This method allows for handling the inherent uncertainty and subjectivity in the decision-making process, providing a robust mechanism to quantify the efficiency of different solutions. The fuzzy entropy approach facilitates a nuanced assessment of the trade-offs and synergies among the identified criteria, enabling stakeholders to make informed, sustainable choices.

To validate the efficacy of our proposed model, we will conduct a results comparative analysis. This study helps policymakers and business leaders in strategic decision-making that aligns with sustainability goals.

SE24

Summit - 344

Technological Change and Social Impact

Contributed Session

Contributed

Chair: Merve Olmez Turan, National Renewable Energy Laboratory, Golden, CO, United States

1 - Cultural Values Impact on Perceptions of Intrusiveness and Legitimacy of AI-Based Managerial controls: A Cross-Country Examination

Akram Khattab, The University of Toledo, Toledo, OH, United States, HEBA ABDEL-RAHIM, Euisung Jung, Paul Hong

Emerging Information and Communication Technologies (ICT) have made employee monitoring a quite prevalent practice. The extensive and intrusive managerial controls which heavily rely on recently emerged artificial intelligence have caused much debate on its legitimacy. Yet, a perspective of AI-Based Managerial controls that has received less attention is how workers perceive Artificial Intelligence (AI) controls. This study aims to exam two research questions: First, what is the general perception of employees to the companies' use of AI controls such as surveillance tools? Second, does the perception of employees to the company's use of AI controls depend on their cultural values? A conceptual model based on Technology Mediated Controls (TMC) theory and Hofstede's cultural value is developed to theorize the effects of AI controls. To test the model, a mixed-source approach is adopted. Individual level data will be collected using a questionnaire with respondents from countries with different cultural values. Country level data will be derived from Hofstede's study (2001). Specifically, the study focuses on the important and pressing issue of perceptions of intrusiveness of AI-based managerial controls in the workplace given employees' national cultural backgrounds. Theoretical and practical implications are presented.

2 - Effect of Interruption On Knowledge Hiding: A Multilevel Study

Rushatey Feroze, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Ruonan Zhao, YuFan Shang

With the advancement of information communication technologies (ICT), job interruptions received through ICT have been widely investigated, while interruptions happened from humans have received less attention. Differentiating these two types of interruptions, this research identifies knowledge hiding as the consequence of human work interruption, and explores the mechanism through which work interruptions impact knowledge hiding based on the conservation of resource theory (COR) from a time management perspective. The data was collected from established countries (over six years) with over 500 employees across several industries. Data were analyzed using Mplus for mediation and cross-level moderation. The results reveal that human work interruption significantly precedes knowledge hiding. In addition, perceived control of time negatively mediates the relationship between work interruptions and knowledge hiding. Also, the team temporal leadership strengthens the negative impact of interruption and perceived control of time and effect of perceived control of time on knowledge hiding. This study seeks to expand the existing research on work interruptions and knowledge hiding by incorporating aspects of team dynamics, particularly temporal leadership, to foster a more comprehensive understanding of knowledge management practices within organizations. Besides, we show a new direction of how strong team temporal leadership not just have positive effect but also negatively affect employees in managing time when faced with human work interruption. These theoretical and practical insights will aid organizations to better manage their work interruption.

3 - A Patent-Based Approach to Analyzing Technological Changes: the Impact of COVID-19 Pandemic on Digital Healthcare Technology

Hyunjin Shin, Ajou University, Suwon-si, Korea, Republic of, Sungjoo Lee

Unexpected social events such as pandemics have radical and far-reaching effects on technological change. Accurately identifying and understanding the ripple effects of these social issues on technological development is essential to strengthening technological responsiveness and establishing long-term development strategies. This study proposes a systematic approach to analyzing technological change made to respond to such events. To achieve this, patents were adopted to extract technological content, based on which pre- and post-event technology trees were developed to be compared for investigating technological changes. The proposed approach was applied to analyzing technological changes in the digital healthcare field before and after the COVID-19 pandemic. This study is expected to contribute to establishing faster and effective technological response strategies, promoting technological development, and securing long-term competitiveness by identifying and interpreting technological changes brought about by social events.

4 - The aggregate effects of firms' IT investments on the structure of a production network and the industry-level prices of intermediate inputs

Fengmei Gong, University of La Verne, La Verne, CA, United States, Byungwan Koh, Barrie Nault

A growing body of literature argues that as firms change their suppliers in response to technological innovations, the structure of their production network is endogenous to technological innovations. Furthermore, it suggests that information technology (IT) has plays a crucial role in technological innovations and in subsequent economic outcomes. Drawing from these insights, we investigate whether firms' IT investments are associated with the structure of a production network and, subsequently, whether the structure of a production network is associated with industry-level prices of intermediate inputs while controlling for the direct relationship between firms' IT investments and the industry-level prices of intermediate inputs. Our measure of the structure of a production network is structural holes. A structural hole is the absence of a direct tie between two alters within an ego network. Using the data constructed from input-output tables, we find that aggregate

IT investments made by firms in an industry are positively associated with structural holes at the industry level. Furthermore, we find that the firms' IT investments are negatively associated with industry-level prices of intermediate inputs not only directly but also indirectly through structural holes at the industry level.

5 - ReEDS Performance Improvement

Merve Olmez Turan, National Renewable Energy Laboratory, Golden, CO, United States, Atharv Bhosekar, Maxwell Brown, Adam Christensen, Wesley Cole

The Regional Energy Deployment System (ReEDS) is an open-source, spatially explicit, long-term capacity expansion model for the bulk electric power system of the contiguous United States, encompassing multiple scenarios with technological and political assumptions (see https://github.com/NREL/ReEDS-2.0). With the increased needs for capabilities, higher temporal and spatial resolutions to model the evolution of the power system with modern technologies and low-carbon pathways, ReEDS' model solution times have increased significantly from 4-6 hours in 2018 to 18-48+ hours in 2023 . Also, the model size for commonly-run ReEDS scenarios reached 22 and 28 million equations and variables, respectively. These runtimes can be especially challenging under certain scenario settings (e.g., very high temporal or spatial resolution) or with limited computational power. In this presentation, we will discuss several methods we used to improve model runtime, including data preparation, model modification, and solver tuning. The implementation of these methods shrank the model size to 7.2 and 7.3 million equations and variables, respectively. Furthermore, this led to a 77% reduction in the model's run time for commonly-run ReEDS scenarios. We will discuss the process of identifying areas for solve time improvements and how the specific enhancements for the ReEDS model might be applied to other similar large-scale models.

SE25

Summit - 345

Information Seeking Behavior in Decision Analysis

Invited Session

Decision Analysis Society

Chair: Zack Zhu, University of British Columbia, Surrey, BC, Canada

1 - Learning from Consideration Sets

Canan Ulu, Georgetown University, Washington, DC, United States, Bharadwaj Kadiyala, Dorothee Honhon

We consider the optimal assortment under the Random Consideration Set model when the firm can learn about consumers' product consideration probabilities and show that it depends on the consumers' preference order and the product informativeness order. When these orders are identical, the optimal assortment is a popular set.

2 - Search in the Dark

Sasa Zorc, University of Virginia, Darden School of Business, Charlottesville, VA, United States, Manel Baucells

The classic sequential search problem rewards the decision maker with the highest sampled value, minus a cost per sample. If the sampling distribution is unknown, then a Bayesian decision maker faces a complex balance between exploration and exploitation. We solve the stopping problem of sampling from a Normal distribution with unknown mean and unknown variance and a conjugate prior, a longstanding open problem. The optimal stopping region may be empty (it may be optimal to continue the search regardless of the offer one receives, especially at the early stages), or it may consist of one or two bounded intervals. While a single reservation price cannot describe the optimal rule, we do find an optimal index policy taking the form of a standardized reservation rule: stop if and only if the standardized value of the current best exceeds a threshold that depends on the standardized search cost. We also provide an algorithm to compute the index function, producing a practical way to implement the optimal stopping rule for any given prior, sampling history, and sampling horizon.

3 - Decision Analytic Explanations for Information Avoidance and Willful Ignorance

Zack Zhu, University of British Columbia, Surrey, BC, Canada, Steven Shechter, Woonghee Huh

In formalizing information acquisition decisions as combined instrumental hedonic utility maximization problems, we develop insights on the degree to which individuals seek vs. avoid information in different settings. These settings include both decisions that affect primarily oneself (e.g., test results leading to disease treatment decisions) as well as those involving self vs. others tradeoffs (e.g., quarantining after a positive Covid test). We bring a behavioral lens to our model, which allows for a negative VOI, and thus rigorously explains why for some people, "ignorance is bliss."

SE26

Summit - 346

Convex and Stochastic Optimization

Contributed Session

Contributed

Chair: Fatih Cengil, University of Arkansas, Fayetteville, AR, United States

1 - Training Neural Networks with User-Defined Dimension-Specific Convexity

Hyunho Kim, Sungkyunkwan University, Suwon, Korea, Republic of, Jong-Seok Lee

Convex functions play an important role in many areas of mathematics, such as optimization and probability theory, due to their convenient properties. This study proposes a novel neural network architecture that approximates convex functions over a user-defined specific subset of features. To estimate such a dimension-specific convex function, we perform data augmentation through symmetrization, utilizing all possible symmetrical combinations for the dimensions requiring convex constraints. Additionally, we employ a partially connected neural network with exponentiated weights to preserve convexity and monotonicity. The proposed method integrates these convex monotonic neural networks to make a final prediction. This approach ensures convexity for user-defined dimensions while providing highly expressive function estimation based on neural networks. We validated the effectiveness of the proposed method through various experiments, showing that it is both effective and practically applicable across different areas.

2 - Approximate IP Models for Lot Sizing and Cardinality Constraints in Portfolio Optimization

Raphael Hauser, University of Oxford, Oxford, United Kingdom

Portfolio optimization with lot sizing or cardinality constraints leads to quadratic mixed integer programming problems that can usually only be solved in limited dimension. While thresholding and rounding approaches are fast and work well most of the time, they can be substantially suboptimal in certain circumstances. We discuss mixed linear integer programming approximations that are scalable in practice and perform favorably in comparison with rounding on practical portfolio problems.

3 - EXPRESS: Automatic Convexity Detection and Transformation to Conic Forms for Nonlinear Optimization

Qiuwei Li, Alibaba US, DAMO Academy, Bellevue, WA, United States, Liang Zhao, Minghan Yang, Wotao Yin, Kuo-Ling Huang

Nonlinear optimization problems are often more effectively solved when converted into conic optimization forms, which leverage advanced conic solvers. These solvers generally require prior transformations into canonical forms, a step that is often laborious. To streamline this process, we introduce EXPRESS, an innovative optimization toolkit designed to automatically detect the convexity of nonlinear functions and transform these problems into their equivalent conic forms. When direct transformation is infeasible, EXPRESS provides symbolic analysis of function convexity to guide the choice of a solver. We validate the effectiveness of EXPRESS with experimental results on the MINLPLib and CUTEST datasets, demonstrating its proficiency in simplifying and accelerating nonlinear optimization processes.

4 - Perturbed Gradient Descent using Tempered Linear Fractional Stable Noise

Farzad Sabzikar, Iowa State University, Ames, IA, United States, Mohammad Fili, Guiping Hu

Navigating the complex landscapes of non-convex optimization presents significant challenges in machine learning, particularly when traditional gradient descent methods with Gaussian assumptions fail to escape local minima or reach more favorable, flat regions. To tackle these challenges, our approach perturbs the gradient descent by introducing tempered linear fractional stable noise (TLFSN), derived from the increments of tempered fractional stable motion (TLFSM), significantly enhancing the optimizer's ability to overcome these limitations. Our methodology is distinguished by incorporating a tempering technique that controls the magnitude of noise-induced jumps. This innovative approach manages the impact of heavy-tailed, correlated noise and marks a pioneering advancement in optimization. By integrating TLFSN and TLFSM, we transform how gradient descent navigates non-convex functions, offering a structured and effective exploration strategy. Theoretically, we establish a comprehensive framework to analyze the effects of tempering on the noise characteristics and the resulting behavior of optimizers. We derive critical bounds that elucidate the influence of heavy-tailed noise on optimization dynamics, providing profound insights into stochastic processes. Empirical validations of our approach demonstrate significant improvements in optimizer performance, showing deeper landscape exploration and more efficient convergence. These findings highlight the transformative potential of our TLFSN method, setting new standards for optimization practices in machine learning.

5 - On the convex hull of bilinearly constrained set

Hyun-Ju Oh, Clemson University, Clemson, SC, United States, Margaret Wiecek, Boshi Yang

In this talk, we address convexification of sets defined by functions with multiple bilinear terms. We present conditions for the sets to have polyhedral convex hulls and for obtaining the closed form of defining constraints for the convex hull. By exploiting these results, we propose polyhedral approximations of the convex hulls. We present computational experiments to compare our approach with the classical methods in terms of tightness and convergence.

6 - Learning to Accelerate Tightening of Convex Relaxations of the AC Optimal Power Flow Problem

Fatih Cengil, University of Arkansas, Fayetteville, AR, United States, Harsha Nagarajan, Russell Bent, Sandra Eksioglu, Burak Eksioglu

We propose a machine learning (ML)-based method aimed at accelerating convergence to global solutions for the Alternating Current Optimal Power Flow (AC-OPF) problem. Our method enhances the efficiency of the optimality-based bound tightening (OBBT) algorithm, known in the literature for effectively tightening variable bounds in the non-convex AC-OPF problem. While the OBBT algorithm provides nearly global optimum solutions through tight convex relaxations, its inherent computational burden persists, even for medium-scaled power networks. The proposed novel ML-based policy, integrated into the OBBT algorithm, replaces the exhaustive OBBT without compromising optimality guarantees. This policy dynamically selects a subset of variables whose sequential bound tightening contributes to tighten the convex relaxation of the AC-OPF problem. To this end, we leverage historical data to learn a correlation between the load profiles and variable subsets via their rankings. Our policy, coupled with the parallel implementation of the OBBT algorithm, facilitates the discovery of near-global optimal solutions in significantly reduced computation times. We demonstrate this through detailed numerical experiments on medium to large-scale instances with up to 3,375 buses. Across the held-out set of instances, we observe up to 20x speed-up in run times of the ML-accelerated OBBT algorithm. To our knowledge, this is the first ML-based approach to demonstrate such large speed-ups for tightening convex relaxations on realistic large-scale power grid instances.

SE27

Summit - 347

Impact of AI on New Product Development: A Panel Discussion

Panel Session

Technology, Innovation Management and Entrepreneurship

Co-Chair: Sreekumar Bhaskaran, SMU-Cox School of Business, Dallas, TX, United States

Co-Chair: Karthik Ramachandran, Georgia Institute of Technology, Atlanta, GA, United States

1 - Panelist

Nitin Joglekar, Boston University Questrom School of Business, Boston, MA, United States

2 - Panelist

Anjana Susarla, Michigan State University, Okemos, MI, United States

3 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

4 - Panelist

Nitin Mayande, Tellagence, Naperville, IL, United States

na

5 - Panelist

Alf Wang, Amazon, Bellevue, WA, United States

6 - Moderator Panelist

Sreekumar Bhaskaran, SMU-Cox School of Business, Dallas, TX, United States

SE28

Summit - 348

Recent Development in Sustainable and Socially Responsible Operations

Invited Session

MSOM: Sustainable Operations

Chair: Anyan Qi, The University of Texas at Dallas, Richardson, TX, United States

1 - Area Conditions and Positive Incentives: Engaging Local Communities to Protect Forests

Xavier Warnes, Indiana University, Bloomington, IN, United States, Joann de Zegher, Dan Iancu, Erica Plambeck

Agricultural production is a dominant driver of tropical deforestation, causing alarming CO2 emissions and loss of biodiversity and ecosystem services. Complex land tenure systems coupled with low enforcement of conservation laws enable farmers to convert tropical forests into productive land for a better income. We propose positive incentives with area conditions for local communities to prevent this deforestation while increasing farmer welfare. We model the deforestation decisions of the community using a cooperative game in partition correspondence form and show how these area-conditional incentives can leverage the existing social capital in agricultural communities. Finally, we demonstrate the effectiveness of these conditions by using data from the Indonesian context.

2 - Managing Online Service Platforms: Service Provision, Contract Design, and Customer Review

Xin Fang, Singapore Management University, Singapore, Singapore, Linqiu Li, Qiyuan Deng, Yun Fong Lim

This paper studies the service provision, contract design, and customer review of online service platforms in the presence of asymmetric quality information. We consider an accommodation-sharing platform on which a landlord lists a house for booking. The house can be of high or low quality, which is unknown to customers. To increase customers' willingness to pay, the landlord can provide value-added service. For every successful booking, the platform charges a commission fee. By analyzing the equilibrium decisions of the three parties, we find that the high-quality landlord can adopt a higher service level and price to signal quality. A costless separating equilibrium emerges if the high-quality landlord has a significant advantage in service cost. To maximize its profit, the platform's optimal commission fraction changes based on the service cost factor and the reservation profit difference between the high and low-quality landlords. We further develop a two-period model in which the second period's customer may learn the house quality from the review left by the first period's customer. We find that the emergence of review makes a costless separating equilibrium achievable even if the two types of landlords have similar service costs. The high-quality landlord benefits from a higher review access rate, but interestingly, a moderate review access rate and service cost factor may maximize social welfare.

3 - Reusable Packaging with Consumer Incentives

Yunlong Peng, University of Warwick, Coventry, United Kingdom, Fei Gao, Jian Chen

An increasing number of consumers have begun using their own reusable packaging to mitigate disposable packaging waste. However, the considerable inconvenience associated with using reusable packaging has constrained consumer adoption. This paper explores two strategies implemented by profit-maximizing firms to encourage the adoption of reusable packaging: (i) offering incentives, including monetary and pro-environmental incentives, and (ii) providing firm-owned reusable packaging to consumers. We study the effectiveness of these strategies and examine their environmental implications.

4 - Combating Excessive Overtime in Global Supply Chains

Chuanya Jiao, University of Science and Technology of China, Hefei, China, People's Republic of, Anyan Qi, Jiayu Chen

Suppliers operating in developing economies may resort to compelling their workforce to engage in excessive overtime, resulting in severe physical and mental health issues for workers and the potential for significant damage to the brand image of multinational enterprises (MNEs) if these practices are exposed to the public. In this paper, we develop a game-theoretic model of a dyadic supply chain to analyze a manufacturer's operational strategies to combat the use of excessive overtime by a supplier. These strategies encompass a stick strategy of auditing the supplier's practice (i.e., the auditing strategy) and carrot supplier-development strategies of subsidizing the supplier's capacity improvement initiative (i.e., the capacity-subsidy strategy) and cross training the supplier's workers to increase their versatility (i.e., the cross-training strategy). When both capacity subsidy and auditing are viable, interestingly, capacity subsidy may be a complement for auditing, contrary to the naive belief that the strategies are always substitutes in combating excessive overtime. Compared to the case when auditing is the sole viable strategy, we find that capacity subsidy may backfire, increasing excessive overtime and decreasing social welfare, when capacity subsidy and auditing are substitutes. Furthermore, the capacity subsidy could lead to social welfare that is even higher than that in a centralized supply-chain benchmark without the capacity subsidy. In situations where both cross-training and auditing are viable, cross training may also be a complement for auditing, driven by the enhanced flexibility of the labor force. However, cross-training may also backfire, resulting in increased excessive overtime and decreased social welfare.

5 - Index-Based Yield Protection for Smallholder Farmers

Can Zhang, Duke University, Durham, NC, United States, Kehan Lu, Jing-Sheng Song

Government subsidies are common in the agricultural sector to protect farmers from unexpected losses. Two major forms of agricultural subsidies are price protection, under which farmers are subsidized when the market price is low, and yield protection, under which farmers are subsidized when the crop yield is low. While price protection is popular in both developed and emerging economies, implementing yield protection in emerging economies is challenging due to the high costs of yield assessment for small farms. This research examines the design of a recently emerged index-based yield protection policy, which triggers subsidies when a pre-determined index (e.g., rainfall) predicts a low yield, thereby avoiding costly yield assessment. Our analysis highlights how the (in)accuracy of the index can lead to nuanced trade-offs and novel insights, and provides recommendations for the design of index-based subsides.

SE29

Summit - 420

Infrastructure Network Resilience

Invited Session

OPT: Network Optimization

Chair: Pelin Kesrit, Texas A&M University, College Station, TX, United States

Co-Chair: Bahar Çavdar, Rensselaer Polytechnic Institute, Troy, NY, United States

1 - Simple v/s non-simple repair policies for networks under simultaneous failures of its components

Guido Lagos, Universidad Adolfo Ibáñez, Viña del Mar, Chile, Cristóbal Reyes, Héctor Olivero, Jorge Navarro

In this work, we study repair policies for reliability networks that are subject to simultaneous failures of its components. Indeed, we assume that “shocks” arrive in a continuous-time Markovian fashion, each shock hits a subset of the components, and once a component is hit it stops working. Depending on the configuration of working and failed components, the network can be either working or non-working. Furthermore, a failed component can be repaired at a component-repair cost, and also the whole network can be repaired once it fails, at a (higher) network-repair cost. Hence, a trade-off arises between saving the component-repair cost by letting them fail, and saving the network-repair cost by repairing components so the network has less chance of failing. We study several repair policies in this setting. Indeed, we consider a family of simple but interpretable repair policies where all failed components are repaired when r or more components have failed, and derive exact expressions for the expected network failure time and for the long-term average cost. Then we consider more involved policies derived using Markov Decision Processes, and compare their performance to the simple repair policies. We study conditions under which one policy is better than the other.

2 - Logistics of Infrastructure Network Restoration under Incomplete Inforamtion

Bahar Çavdar, Rensselaer Polytechnic Institute, Troy, NY, United States, Joseph Geunes, Subhojit Biswas

In this paper we study repair crew routing problem for infrastructure network restoration under incomplete information. We consider a disrupted infrastructure network where the dispatchers know the locations of service losses but not necessarily the actual fault locations. We model this problem as a finite horizon Markov decision process and develop efficient solution methods using reinforcement learning methods. Our solution method uses structural results and state-space aggregation to address the time-sensitive nature of such decisions.

3 - Fortifying Distribution Network Nodes Subject to Disruptions

Pelin Kesrit, Texas A&M University, College Station, TX, United States, Bahar Çavdar, Joseph Geunes

We consider a distribution network for delivering a natural resource or physical good to a set of nodes, each of which serves a set of customers, in which flow disruptions may occur at one or more nodes. Each node receives flow through a path from one or more source nodes, implying that a node experiences a disruption if a disruption occurs at one or more nodes on each path from a source node. All nodes in the network are subject to a future disturbance of an uncertain degree of severity, and we assume we can quantify the degree of severity, and that it follows some well-defined probability distribution. For each node in the network, we wish to determine a fortification level that enables the node to withstand a disturbance up to some level of severity. The cost of fortification is nondecreasing in the maximum-severity fortification level chosen, and we wish to maximize the expected number of customers who do not experience a disruption following the occurrence of a disturbance, given a limited fortification budget. We formulate this problem as a mathematical program, characterize useful properties of optimal solutions, and provide methods for determining optimal fortification levels under various assumptions on the probability distribution of the disturbance severity and the network structure.

4 - Balanced Design of Important Components in Multi-Commodity Infrastructure Networks

Kash Barker, University of Oklahoma, Norman, OK, United States, Andres Gonzalez, Himadri Sen Gupta

In multi-commodity networks, determining which components to focus on for reducing the impact of a network disruption can be done by evaluating a metric of importance of each component to the network’s operation. On the other hand, having components in a multi-commodity network that are much more critical than others may expose the network to potentially significant disruptions as well as targeted attacks. To address this issue, we propose a novel bi-objective optimization formulation to determine optimal actions for balancing the importance of components, illustrating with a multi-commodity infrastructure network example.

SE30

Summit - 421

Advances in Conic Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Pouya Sampourmahani, Lehigh University, Bethlehem, PA, United States

1 - Interior Point Methods for Nearly Linear Time Algorithms

Aaron Sidford, Stanford University, Stanford, CA, United States

In this talk I will survey recent advances in designing and applying interior point methods to obtain improved running times for foundational continuous and combinatorial optimization problems. I will discuss how recent interior point methods can be leveraged to reduce linear programming, maximum cardinality bipartite matching and more to data structures problems that how these problems can be solved efficiently with techniques from randomized numerical linear algebra and graph theory. In particular, I will highlight joint work that showed linear programs with d variables and n constraints can be solved with high probability to high precision in time O~(nd + poly(d)) and touch upon its implications for solving combinatorial optimization problems.

2 - Faster LP Solvers via Robust Interior Point Methods

Jan van den Brand, Georgia Tech, Atlanta, GA, United States

In this talk, I will outline techniques that led to major improvements in the theoretic performance of linear program solvers. A key element is the development of “Robust Interior Point Methods” designed to handle larger numeric or approximation errors. This allows for sublinear time per iteration via more aggressive variants of Karmarkar' acceleration scheme and cruder approximation from randomized linear algebra techniques. A crucial aspect here is measuring the centrality via softmax-functions, which provide a natural relaxation of wider neighborhoods like those based on L_∞-norms, while offering better analytic properties.

3 - On Quadratic Convergence for Semidefinite Optimization Problems

Pouya Sampourmahani, Lehigh University, Bethlehem, PA, United States, Mohammadhossein Mohammadisiahroudi, Ali Mohammad-Nezhad, Brandon Augustino, Saugata Basu, Tamás Terlaky

Attaining a fast convergence rate for interior point methods (IPMs) for solving semidefinite optimization (SDO) problems has historically depended on assuming non-degeneracy and strict complementarity assumptions. Literature shows that failing strict complementarity results in the convergence rate of IPMs not to accelerate. In this presentation, we discuss two approaches where a fast convergence rate can be attained without such assumptions. The first approach is based on higher-order IPMs applied to reparametrized weighted central paths, and the second one is based on a novel approach called iterative refinement for SDOs. We show that the weak assumption of interior point condition is all you need to achieve superlinear or quadratic convergence rates in each approach, respectively. We present numerical results for the first approach. We further show the second approach can be applied to improve the complexity of inexact classical and quantum solvers for SDO.

SE31

Summit - 422

New Min-max Optimization Algorithms and Their Applications in Machine Learning and AI

Invited Session

OPT: Machine Learning

Chair: Yang Tianbao, Texas A&M University, College Station, TX, United States

Co-Chair: Bokun Wang, Texas A&M University, College Station, TX, 77840, United States

1 - An Optimal Single-Loop Algorithm for Convex Finite-Sum Coupled Compositional Stochastic Optimization

Bokun Wang, Texas A&M University, College Station, TX, United States

Our work revisits a class of convex Finite-Sum Coupled Compositional Stochastic Optimization (cFCCO) problems with many applications, including group distributionally robust optimization (GDRO), reinforcement learning, and learning to rank. To better solve these problems, we introduce a unified family of efficient single-loop primal-dual block-coordinate proximal algorithms, dubbed ALEXR. This algorithm leverages block-coordinate stochastic mirror ascent updates for the dual variable and stochastic proximal gradient descent updates for the primal variable. We establish the convergence rates of ALEXR in both convex and strongly convex cases under smoothness and non-smoothness conditions of involved functions, which not only improve the best rates in previous works on smooth cFCCO problems but also expand the realm of cFCCO for solving more challenging non-smooth problems such as the dual form of GDRO. Finally, we present lower complexity bounds to demonstrate that the convergence rates of ALEXR are optimal among first-order block-coordinate stochastic algorithms for the considered class of cFCCO problems.

2 - Advancing Tractability and Generalization for Smart "Predict-Then-Optimize" via Stochastic Minimax Reformulation

Jiawei Zhang, MIT, Cambridge, MA, United States

In stochastic optimization, certain parameters are often unknown, and contextual information is sometimes used to help estimate these key unknown parameters to solve the optimization problem accurately. This approach is known as contextual optimization. In this talk, I will introduce our algorithm designed to solve contextual optimization problems with computational tractability, optimality, and a generalization guarantee. This is achieved through a minimax reformulation.

3 - Nearly Optimal L-p Risk Minimization

Zhichao Jia, Georgia Institute of Technology, Atlanta, GA, United States, Guanghui Lan, Zhe Zhang

Convex risk measures play a foundational role in the area of stochastic optimization. However, in contrast to risk neutral models, their applications are still limited due to the lack of efficient solution methods. In particular, the mean L-p semi-deviation is a classic risk minimization model, but its solution is highly challenging due to the composition of concave-convex functions and the lack of uniform Lipschitz continuity. In this paper, we discuss some progresses on the design of efficient algorithms for L-p risk minimization, including a novel lifting reformulation to handle the concave-convex composition, and a new stochastic approximation method to handle the non-Lipschitz continuity. We establish an upper bound on the sample complexity associated with this approach and show that this bound is not improvable for L-p risk minimization in general.

4 - Doubly Smoothed Optimistic Gradients: A Universal Recipe for Smooth Minimax Problems

Taoli Zheng, The Chinese University of Hong Kong (CUHK), Hong Kong, Hong Kong, Jiajin Li, Anthony So

Smooth minimax optimization finds widespread applications in machine learning and operation research, but existing algorithms often rely on specific structural conditions like convexity, Kurdyka-Łojasiewicz properties and weak minty type conditions, which can be challenging to verify in practice. More importantly, even convex and nonconvex minimax optimization are typically treated separately. Our work takes the first step to provide a universal recipe to handle all of them through the doubly smoothed optimistic gradient descent ascent method (DS-OGDA), a universally applicable algorithm for smooth minimax problems. With a single set of parameters, DS-OGDA achieves $\mathcal{O}(\epsilon^{-4})$ complexity for both nonconvex-concave and convex-nonconcave optimization problems, and $\mathcal{O}(\epsilon^{-2})$ for convex-concave cases. With additional information at hand, DS-OGDA is also able to achieve the optimal or best-known results under each scenario.

SE32

Summit - 423

Robustness at the Intersection of Optimization, Statistics, and Machine Learning

Invited Session

OPT: Optimization Under Uncertainty

Chair: Peter Zhang, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States

1 - An Adversarially Robust Formulation of Linear Regression with Missing Data

Yue Xing, Michigan State University, East Lansing, MI, United States, Alireza Aghasi, Saeed Ghadimi, Mohammad Javad Feizollahi

We present a robust framework to perform linear regression with missing entries in the features. By considering an elliptical data distribution, and specifically a multivariate normal model, we are able to conditionally formulate a distribution for the missing entries and present a robust framework, which minimizes the worst-case error caused by the uncertainty in the missing data. We show that the proposed formulation, which naturally takes into account the dependency between different variables, ultimately reduces to a convex program, for which we develop a customized and scalable solver. We analyze the consistency and structural behavior of the proposed framework in asymptotic regimes, and present technical discussions to estimate the required input parameters. We complement our analysis with experiments performed on synthetic, semi-synthetic, and real data, and show how the proposed formulation improves the prediction accuracy and robustness, and outperforms the competing techniques.

2 - A Smooth Clipping and Error-Feedback Approach for Distributed Optimization Under Heavy-Tailed Noise

Shuhua Yu, Carnegie Mellon University, Pittsburgh, PA, United States, Dusan Jakovetic, Soummya Kar

Motivated by understanding and analysis of large-scale machine learning under heavy-tailed gradient noise, we study distributed optimization with gradient clipping, i.e., in which certain clipping operators are applied to the gradients or gradient estimates computed from local clients prior to further processing. While vanilla gradient clipping has proven effective in mitigating the impact of heavy-tailed gradient noise in non-distributed setups, it incurs bias that causes convergence issues in heterogeneous distributed settings. To address the inherent bias introduced by gradient clipping, we develop a smoothed clipping operator, and propose a distributed gradient method equipped with an error feedback mechanism, i.e., the clipping operator is applied on the difference between some local gradient estimator and local stochastic gradient. We establish that, for the first time in the strongly convex setting with heavy-tailed gradient noise that may not have finite moments of order greater than one, the proposed distributed gradient method has mean-square error (MSE) that converges at a rate $O(1/t^\iota)$, $\iota \in (0, 0.4)$, where the exponent ι stays bounded away from zero, and is shown to be independent of the existence of higher order gradient noise moments $\alpha > 1$. Numerical experiments on both synthetic and real-world data validate our theoretical findings.

3 - On the Adversarial Robustness of Benjamini Hochberg

Louis Chen, NPS, Monterey, CA, United States

The Benjamini-Hochberg (BH) procedure is widely used to control the false detection rate (FDR) in multiple testing. Applications of this control abound in drug discovery, forensics, anomaly detection, and, in particular, machine learning, ranging from nonparametric outlier detection to out-of-distribution detection and one-class classification methods. Considering this control could be relied upon in critical safety/security contexts, we investigate its adversarial robustness. More precisely, we study under what conditions BH does and does not exhibit adversarial robustness, we present a class of simple and easily implementable adversarial test-perturbation algorithms, and we perform computational experiments. With our algorithms, we demonstrate that there are conditions under which BH's control can be significantly broken with relatively few (even just one) test score perturbation(s), and provide non-asymptotic guarantees on the expected adversarial-adjustment to FDR. Our technical analysis involves a combinatorial reframing of the BH procedure as a ``balls into bins'' process, and drawing a connection to generalized ballot problems to facilitate an information-theoretic approach for deriving non-asymptotic lower bounds.

4 - Finding Pareto Efficient Solutions in Robust Optimization with First-Order Methods

Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang

When designing and using robust optimization models, a key concern is to balance efficiency and robustness. For instance, trading-off between clean performance and adversarial performance in adversarial robust machine learning. Typically, model designers control this trade-off by adjusting either the uncertainty set's shape and radius hyperparameters or the so-called globalized/comprehensive robust optimization penalty hyperparameters. In both cases, finding each solution requires solving a different instance of robust optimization problem (with potentially additional computational steps to arrive at the Pareto frontier). This makes balancing efficiency and robustness computationally challenging in practice. In this paper, we show that an approximate efficiency-robustness Pareto frontier can be obtained efficiently in a single-shot via first-order methods for minimax optimization, where the Pareto frontier corresponds to the trajectory of the first-order methods.

SE33

Summit - 424

Distributional Robustness and Learning

Invited Session

OPT: Optimization Under Uncertainty

Chair: Mohammed Amine Bennouna, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Bahar Taskesen, EPFL, Lausanne, Switzerland

1 - Distributionally Robust Path Integral Control

Grani Adiwena Hanasusanto, University of Illinois Urbana-Champaign, Urbana, IL, United States, Hyuk Park, Duo Zhou, Takashi Tanaka

<div zzyype="">

We address a continuous-time continuous-space stochastic optimal control problem where the controller lacks exact knowledge of the underlying diffusion process. Instead, the controller relies on a finite set of historical disturbance trajectories. To tackle this challenge, we propose a novel approach called Distributionally Robust Path Integral control (DRPI). This method utilizes distributionally robust optimization (DRO) to make the resulting policy robust against the unknown diffusion process. Notably, the DRPI scheme bears resemblance to risk-sensitive control, allowing us to leverage the path integral control (PIC) framework as an efficient solution method. We establish theoretical performance guarantees for the DRPI scheme, which facilitates the determination of an appropriate risk parameter for the risk-sensitive control. Through extensive validation, we demonstrate the effectiveness of our approach and its superiority over risk-neutral and risk-averse PIC policies in scenarios where the true diffusion process is unknown.

2 - Diagnosing Model Performance Under Distribution Shift

Tiffany Cai, Columbia University, New York, NY, United States, Steve Yadlowsky, Hongseok Namkoong

Prediction models can perform poorly when deployed to target distributions different from the training distribution. To understand these operational failure modes, we develop a method, called DIstribution Shift DEcomposition (DISDE), to attribute a drop in performance to different types of distribution shifts. Our approach decomposes the performance drop into terms for 1) an increase in harder but frequently seen examples from training, 2) changes in the relationship between features and outcomes, and 3) poor performance on examples infrequent or unseen during training. These terms are defined by fixing a distribution on X while varying the conditional distribution of Y|X between training and target, or by fixing the conditional distribution of Y|X while varying the distribution on X. In order to do this, we define a hypothetical distribution on X consisting of values common in both training and target, over which it is easy to compare Y|X and thus predictive performance. We estimate performance on this hypothetical distribution via reweighting methods. Empirically, we show how our method can 1) inform potential modeling improvements across distribution shifts for employment prediction on tabular census data, and 2) help to explain why certain domain adaptation methods fail to improve model performance for satellite image classification.

3 - Bayesian Risk-Averse Reinforcement Learning

Yuhao Wang, georgia institute of technology, atlanta, GA, United States, Enlu Zhou

In this work, we study the Bayesian risk-averse formulation in reinforcement learning (RL). We adopt the Bayesian risk Markov Decision Process (BRMDP) to account for the parameter uncertainty of the unknown underlying model, for which we derive the asymptotic normality that characterizes the difference between the Bayesian risk value function and the original value function under true unknown distribution. The result indicates the Bayesian risk-averse approach tends to pessimistically underestimate the original value function and such a gap increases with stronger risk-aversion but decreases with more available data. We then utilize this adaptive property in the setting of online RL as well as the online contextual multi-arm bandits (CMAB), a special case of online RL. We provide two procedures using posterior sampling for both the CMAB problem and the general RL problem, where for CMAB we establish the sub-linear regret bound with the regret defined as either the conventional regret or the Bayesian risk regret. Finally, we run several numerical experiments to test the performance of the proposed algorithms and verify the theoretical argument.

4 - It's All in the Mix: Wasserstein Machine Learning with Mixed Features

Aras Selvi, Imperial College Business School, London, United Kingdom, Reza Belbasi, Wolfram Wiesemann

The recent advent of data-driven and end-to-end decision-making across different areas of operations management has led to an ever closer integration of prediction models from machine learning and optimization models from operations research. A key challenge in this context is the presence of estimation errors in the prediction models, which tend to be amplified by the subsequent optimization model – a phenomenon that is often referred to as the Optimizer's Curse or the Error-Maximization Effect of Optimization. A contemporary approach to combat such estimation errors is offered by distributionally robust problem formulations that consider all data-generating distributions close to the empirical distribution derived from historical samples, where 'closeness' is determined by the Wasserstein distance. While those techniques show significant promise in problems where all input features are continuous, they scale exponentially when binary and/or categorical features are present. This work demonstrates that such mixed-feature problems can indeed be solved in polynomial time. We present a practically efficient algorithm to solve mixed-feature problems, and we compare our method against alternative techniques both theoretically and empirically on standard benchmark instances.

SE34

Summit - 425

Methods for Large-Scale Nonlinear and Stochastic Optimization III

Invited Session

OPT: Nonlinear Optimization

Chair: Shagun Gupta, UT Austin, Austin

Co-Chair: Jiahao Shi, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Raghu Bollapragada, The University of Texas at Austin, Austin, TX, United States

1 - Inexact Sequential Quadratic Programming for Noisy Equality Constrained Optimization with Rank-Deficient Jacobians

Jiahao Shi, University of Michigan, Ann Arbor, MI, United States

We design, analyze and implement an inexact sequential quadratic programming (SQP) methods for solving noisy constrained equality optimization problems with an adaptive step size scheme in the absence of constraint qualifications.

2 - Variance-Reduction Method for Variational Inequality Problems

Afrooz Jalilzadeh, The University of Arizona, Tucson, AZ, United States

In this talk, we address variational inequalities (VI) with a finite-sum structure. We introduce a novel single-loop stochastic variance-reduced algorithm, incorporating the Bregman distance function, and establish an optimal convergence guarantee under a monotone setting. Additionally, we explore a structured class of non-monotone problems that exhibit weak Minty solutions, and analyze the complexity of our proposed method, highlighting a significant improvement over existing approaches. Numerical experiments are presented to demonstrate the performance of our algorithm compared to state-of-the-art methods.

3 - A Parameter-Free Gradient Method for Convex Optimization

Shiqian Ma, Rice University, Houston, TX, United States, Danqing Zhou, Junfeng Yang

We propose AdaBB, an adaptive gradient method based on the Barzilai-Borwein stepsize. The algorithm is line-search-free and parameter-free, and essentially provides a convergent variant of the Barzilai-Borwein method for general unconstrained convex optimization. We analyze the ergodic convergence of the objective function value and the convergence of the iterates for solving general unconstrained convex optimization. Compared with existing works along this line of research, our algorithm gives the best lower bounds on the stepsize and the average of the stepsizes. Moreover, we present an extension of the proposed algorithm for solving composite optimization where the objective function is the summation of a smooth function and a nonsmooth function. Our numerical results also demonstrate very promising potential of the proposed algorithms on some representative examples.

4 - Stochastic Algorithms for Constrained Optimization for Informed Learning

Frank E. Curtis, Lehigh University, Bethlehem, PA, United States, Qi Wang

I will present our work on stochastic Newton-based algorithms for solving constrained continuous optimization problems for informed supervised learning. The learning process is informed through the presence of constraints. Using physics-informed learning test problems, we demonstrate that our methods, which are backed by convergence guarantees, are computationally efficient and advantageous in practice. Of particular interest is our newly proposed projection-based variant of the Adam diagonal scaling scheme.

SE35

Summit - 427

Labor Planning at Amazon: Handling Uncertainty

Invited Session

Transportation Science and Logistics (TSL)

Chair: Martin Savelsbergh, Amazon Australia, Darlington, Australia

1 - Workforce Composition: Fixed of Flexible Labor?

Saba Neyshabouri, Amazon, New York, NY, United States

Amazon's Last Mile operations has access to different labor types to address its labor needs for both under-the-roof and on-the-road use cases. These labor types are broadly categorized into fixed and flexible labor pools. Each category offers its specific characteristics regarding cost, reliability, flexibility, and productivity. Coupled with business specific requirements, the existing trade-offs between cost and performance makes the prescription of the optimal workforce composition a non-trivial task.
In this talk, we present simple flow-type models that aim to address the optimal workforce optimization question by accounting for the

interactions between above mentioned characteristics. Our results show the benefits of simplified models to help stakeholders to develop a deeper understanding of the trade-offs and thus develop confidence in more sophisticated models.

2 - Labor Pooling: Nodal Planning

Hadi Panahi, Amazon, Bellevue, WA, United States, Thomas Fillebeen, Roman Levkin

Labor planning is challenging due to demand volatility and supply reliability. Traditionally sites rely on their own labor pool to fulfill demand. However, labor imbalance in geographically neighboring sites can be an opportunity to reduce over and under-staffing by leveraging their shared labor pool. In 2023, Amazon conducted the 'nodal planning' pilot in which adjacent sites, namely 'nodes', shared their labor pool to explore the potential benefits of cross site labor sharing. We leveraged a capacity simulation to estimate the benefits of nodal labor sharing. In addition, we propose an approach for nodal posting policy.

3 - Capacity-Aware Flexible Shift Planning for Under-The-Roof Operations at Amazon

Ramon Auad, Amazon, Santiago, Chile, Thomas Fillebeen, Roman Levkin

Amazon's Last Mile operations have traditionally relied on Fixed associates, who work recurring weekly schedules to meet labor demands. These associates can flex their hours up or down through voluntary extra time (VET) or voluntary time off, respectively. However, the flexibility provided by Fixed associates falls short in accommodating the variability in demand. To address this challenge, Amazon introduced Ready associates, a program that offers its members on-demand voluntary schedules, providing greater flexibility. While flexible labor (VET and Ready) helps reduce overstaffing costs, each option has its own business requirements, costs, capacities, and associate preferences.
In this talk, we present two mixed-integer programming models aimed at improving the overall effectiveness of flexible shift planning. A first long-term planning model utilizes historical data to estimate the distribution of capacity that Ready associates can offer throughout the week's hours; this model influences downstream hiring decisions and better aligns Fixed schedules with Ready associate preferences. The second short-term model jointly optimizes the planning of on-demand VET and Ready shifts, subject to lever-specific capacity constraints, business rules, and the hired Fixed capacity; these planned shifts are then offered to flexible associates to fill any remaining labor gaps. We show preliminary results that showcase the effectiveness of this approach.

4 - Intraday Under-the-Roof Staff Allocation for the Ultra-Fast Delivery at Amazon

Zeynep Sargut, Amazon, Bradenton, FL, United States, Ramon Auad

Same-Day Delivery is an ultra-fast business that offers delivery only a few hours after an order is placed for a selection of time-sensitive and essential items. Finding the best Under-the-Roof staff allocation across multiple processes just before the time of operations is a highly complex problem. Incorrect staff allocation may result in situations where we stop receiving orders for a given fast delivery window, leading to late deliveries or low staff utilization. The goal is to generate a robust plan that respects complex operational constraints and minimizes the risk of being out of capacity while maintaining acceptable staff utilization. For example, plans that involve frequent labor movements are not easy to execute and result in unproductive time. In this talk, we will introduce an optimization-based approach to support intraday staff allocation decisions.

SE36

Summit - 428

Design and Optimization of Societally-aware Mobility Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Devansh Jalota, Stanford University, Stanford, CA, United States

Co-Chair: Matthew Tsao, Lyft, San Francisco, CA, United States

Co-Chair: Gioele Zardini, Stanford University, Stanford, United States

1 - Designing and Operating Large-Scale Multi-Modal Transit Systems

Samitha Samaranayake, Cornell University, Ithaca, NY, United States

Affordable, equitable and efficient access to personal mobility is a fundamental societal need---with broad implications for personal well-being, economic mobility, education, and public health. Achieving this goal with a transportation system that is sustainable and scalable is unlikely to be possible without a significant mass-transit component supported by other modes. This raises many interesting questions on how new (and old) technologies can be integrated with mass transit to improve personal mobility and related societal goals (e.g., sustainability and equity). This talk will focus on some algorithmic and practical questions related to designing and operating integrated multi-modal transit systems. We will discuss the problem setting at a high-level and talk about some specific challenges/solutions in the context of bus line-planning in the multi-modal setting.

2 - Accessibility Insufficiency Minimizing Mobility Hub Design Using Stochastic Assignment Games with Recursive Logit

Joseph Chow, New York University, Brooklyn, NY, United States, Bingqing Liu, Xiyuan Ren

Mobility hubs are facilities that integrate multiple mobility services together, with the intent of not only facilitating transfers, but also providing support for fleets with idling/parking, charging/refueling, and maintenance as necessary. There is a lack of understanding of how mobility hubs can serve as control levers for public agencies to engage with mobility providers. In this study, we customize the platform framework from our earlier stochastic assignment game work to specifically design for mobility hubs in which subsidies are considered to incentivize mobility providers to serve. We solve a Stackelberg game using recursive logit to determine lower level flows and upper level prices to minimize accessibility insufficiency for equity. Measures are shown for quantifying the welfare and equity impacts of a mobility hub siting decision.

3 - A Unified Toll Lane Framework for Autonomous and High-Occupancy Vehicles in Interactive Mixed Autonomy

Ruolin Li, University of Southern California, Los Angeles, CA, United States, Philip Brown, Roberto Horowitz

We propose a toll lane framework for autonomous vehicles and high-occupancy vehicles on shared freeways where autonomous vehicles, capable of maintaining shorter headways, enhance traffic throughput. We designate a restricted, tolled lane, free for autonomous high-occupancy vehicles to boost social mobility, while other vehicles need to pay a toll for access. Assuming vehicles aim to minimize individual travel costs, we explore lane choice equilibria and rank vehicle types by their mobility-enhancing potential—introducing the "mobility degree". Using numerical examples, we address design challenges spanning optimal toll, occupancy threshold, and lane policy, demonstrating the framework's potential to harmonize high-occupancy and autonomous vehicle integration. We develop an algorithm for rational toll differentiation across vehicle types to efficiently reduce total commuter delay and assess the impact of potential vehicle misbehavior, like unauthorized toll lane use at reduced rates. Results show that self-interest-driven behaviors help mitigate impacts of moderate toll non-compliance, highlighting the framework's resilience. This study marks the first systematic exploration of a toll lane framework that facilitates the coexistence of autonomous and high-occupancy vehicles, providing insights for enhancing traffic management and integrating autonomous vehicles into existing transport infrastructure.

4 - Bottleneck Reservation Games: Stable Equilibrium and Its Inefficiency Bounds

Xi Lin, University of Michigan, Ann Arbor, MI, United States, Yafeng Yin

Connected and automated driving technologies could potentially mitigate congestion by letting drivers reserve time slots during peak hours. This study investigates how well this reservation system would work. Specifically, we explored the notion of stable equilibrium resulting from this type of games, and established theoretical bounds on the inefficiency of worst-case stable equilibrium under a range of general and specific scenarios.

5 - Quantifying Uncertainty: Advancing Robustness, Reliability, and Fairness in AI-Driven Transportation Demand Modeling

Dingyi Zhuang, MIT, Canbridge, MA, United States

This research addresses the limitations of traditional AI-driven transportation demand models, which predominantly focus on accuracy at the expense of robustness, reliability, and fairness. Robustness ensures models can handle unexpected data variations, reliability guarantees consistent performance across different conditions, and fairness ensures equitable service provision across diverse user groups. The lack of these attributes can lead to transportation systems that are unprepared for emergencies, inefficient resource allocation, and systemic biases that exacerbate social inequalities. To address these challenges, the study integrates Uncertainty Quantification (UQ) into AI-driven transportation demand models. UQ measures potential inaccuracies in model predictions, enhancing robustness by revealing performance under various scenarios, contributing to reliability by identifying prediction stability, and promoting fairness by highlighting disparities across different groups. The proposed framework not only achieves high accuracy but also emphasizes robustness, reliability, and fairness, thereby improving the overall efficiency and equity of transportation systems. This research includes the development of theoretical and practical methodologies for UQ and assesses the social impacts to ensure equitable transportation planning.

SE37

Summit - 429

Robust and Fair Optimization Problems in Transportation and Supply Chain

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Yidi Miao, Carnegie Mellon University, 5000 Forbes Ave, Hamburg Hall, Pittsburgh, PA, 15213, United States

Co-Chair: Peter Zhang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Retail on Autonomous Wheels: a TIME-Sensitive Traveling Salesman Problem

Zhuolun Dong, University of Texas at Austin, Austin, TX, United States

The rapid advancement in self-driving technology has brought mobile retail into the spotlight. Motivated by its tremendous potential, we present a model, theory, and insights into the operations of the business model, where a wheeled store traverses across a service region, selectively seeks locations in which to perch, and fulfills local customer demand. The goal of the retailer is to balance between perching and migrating to maximize the total revenue during a service time window. To dispatch such a store, the retailer needs to consider the following time-sensitive factors: (i) How busy the service is, (ii) how soon customers come to the store, and (iii) how fast the product freshness decays. We propose and analyze a model that incorporates these factors into the classical traveling salesman problem (TSP), which we call the "time-sensitive TSP" (TSTSP). Using the continuous approximation approach, we derive the order of the optimal revenue, as well as the upper and lower bounds in the asymptotic regime as demands increase to infinity. The tightness of bounds is validated. Our analytical and numerical analysis demonstrates that store mobility creates value by bringing proximity to customers and by taking advantage of flexible repositioning to reach high-profit areas. In particular, the store opts to migrate more frequently during the initial period of the business hours than later on. The store also shifts more time to perching if customer response time becomes longer. These findings demonstrate the potential of mobile retail to emerge as a competitive new retail form.

2 - Inventory Management for Promotional Sales Under Limited Information

Doran Wood, Southern Methodist University, Dallas, TX, United States, Sila Cetinkaya, Eojin Han

We consider a retailer operating in a promotional sales setting and evaluate both risk-neutral and risk-averse inventory policies under limited information. Our primary goal is to analytically characterize the worst-case demand distribution and identify optimal operational policies in closed-form. In addition to analytical takeaways and managerial insights, numerical experimentation is conducted across varying demand distributions to verify our results and to show the advantages and disadvantages between the various policies and modeling approaches.

3 - Fair Optimization of Mobility Subsidies in Multimodal Transportation Networks

Lindsay Graff, Carnegie Mellon University, PITTSBURGH, PA, United States

Several American cities recently piloted Universal Basic Mobility programs with the intention of removing the monetary cost barrier to transportation. In these pilots, the government distributed a mobility subsidy to eligible residents to be used on various mobility services: public transit, transportation network company, bikeshare, and shared scooters. Despite research showing that accessibility indicators vary across space, all participants received the same monetary benefit regardless of residence. The consequence of this policy is that some residents are required to use more of their mobility funds to reach the same destination in the same amount of time, relative to other residents. This highlights a potential limitation in the current approach to program design. In this work, we address this shortcoming and present a method for the optimization of mobility subsidies, where the objective is to maximize job accessibility for eligible subsidy recipients. Our method explicitly accounts for fairness by ensuring that recipients can each reach a similar number of jobs using the city's multimodal transportation network. This is accomplished by differentially allocating the program budget based on a recipient's residence to account for the fact that transit service and micromobility infrastructure are location dependent. We illustrate the method using open-source data from Pittsburgh, PA, one of the cities to implement a Universal Basic Mobility program. Our results show that mobility subsides are a promising instrument to improve access to jobs and alleviate transportation discrepancies.

4 - Transportation Service Region Design Under Distributional Uncertainty

Yidi Miao, Carnegie Mellon University, Pittsburgh, PA, United States

We propose a novel framework for the design of dedicated transportation service regions, aimed at optimizing operations cost through the partitioning of the service region. By leveraging the accumulated operational data over time, our framework assumes that dedicated regions for vehicles can be effectively managed and optimized. To address the inherent ambiguity in demand realization, we adopt a data-driven distributionally robust optimization approach, preventing decisions from out-of-sample effects. Our primary objective is to partition the entire service area into several contiguous and independent districts, while minimizing overall operational costs. This is achieved through a discrete approximation of the worst-case distribution within the ambiguity set, coupled with contiguity constraints to ensure practical implementation. The proposed framework is robust and adaptable, making it applicable to similar transportation and logistics problems.

5 - Optimal Allocation and Route Design for Station-Based Drone Inspection of Large-Scale Facilities

Lei Cai, wuhan university, wuhan, China, People's Republic of

The utilization of drones to conduct inspections on industrial electricity facilities, such as large-sized wind turbines, has recently received significant attention, mainly due to its potential to enhance inspection efficiency and save maintenance costs. Motivated by the advantages of drones for facility inspection, we present a novel station-based drone inspection problem (SDIP) for large-scale facilities. The objective of SDIP is to determine the locations of multiple homogeneous automatic battery swap stations (ABSSs) equipped with drones, assign facility inspection tasks to the ABSSs with operation duration constraints, and design drone inspection routes with battery capacity constraints, such that minimize the sum of fixed ABSS costs and drone travel costs. The SDIP can be regarded as a variant of the location-routing problem, which is NP-hard and difficult to solve optimally. To obtain the optimal solution of SDIP efficiently, we firstly formulate this problem into an arc based formulation and a route based formulation, and then develop a logic-based Benders decomposition (LBBD) algorithm to solve it. The numerical results on five scales of randomly generated instances validate the effectiveness of the LBBD algorithm. Specifically, the LBBD can solve all small- and middle-sized instances, and seven out of ten large-sized instances in 1000 seconds. Furthermore, we conduct a sensitivity analysis by varying the attributes of ABSSs and drones, and provide valuable managerial insights for large-scale facility inspection.

SE38

Summit - 430

AAS Student Presentation Competition III

Award Session

Aviation Applications

Chair: Wayne Ng, Singapore University of Technology and Design, Singapore, Singapore

SE39

Summit - 431

RAS Problem Solving Competition

Invited Session

Railway Applications

Chair: Xuesong (Simon) Zhou, ASU, 11235 E Beryl Ave, Tempe, AZ, 85259, United States

1 - Yard Optimization Problem Definition

David Hunt, Oliver Wyman, Princeton, NJ, United States, Marc Meketon

The RAS problem solving competition participants will present the details of their yard optimization problem solutions, including the problem definition, specific problem sets to be analyzed, solution approach and results.

SE40

Summit - 432

Intelligent Suggestions: Empowering Human Decision Dynamics

Invited Session

Behavioral Operations Management

Co-Chair: Lijia Tan, Eindhoven University of Technology, Helmond, Netherlands

1 - Demand Forecast for Product Selection with Signals: An Experimental Study

Wanshan Zhu, Renmin University of China, Beijing, China, People's Republic of, Manqi Li, Juan Li

In live streaming e-commerce, selecting the right product to be live-streamed is a crucial decision that relies on available signals indicating the products' market popularity. However, human bias in decision makers' interpreting these signals may result in a product selection mistake, where a less profitable product is chosen for live-streaming. To address this issue, we develop a behavioral theory and characterize the necessary and sufficient conditions of signal strength and quantity for product selection mistakes. By conducting experiments, we estimate the parameters of human bias and validate the conditions for such mistakes. The experiments validate our theoretical findings, showing that when faced with two products, human subjects may select a less profitable product due to either overestimation or underestimation. Specifically, these mistakes occur when one product has higher signal strength but lower signal quantity compared to the other, or vice versa. This highlights the importance of considering both signal strength and quantity for correct product selection in live streaming e-commerce.

2 - Algorithm Reliance in Queueing Systems

Stephen Leider, University of Michigan, Ann Arbor, MI, United States, Clare Snyder, Samantha Keppler

Decision-support algorithms promise potential gains to service efficiency and quality, but the realized value of these tools depends on whether and how workers actually use them to serve their customers. How do workers within queueing systems use algorithms to manage system loads? With a laboratory experiment varying system load and algorithm quality, we find people are sensitive to both dimensions: they are more willing to follow their algorithm's advice when its quality is superior or their system loads are high. However, algorithm use does not necessarily lead to better performance, in particular, faster throughput times. In fact, subjects' throughput times only improve over the no-algorithm baseline when the system load is high *and* algorithm quality is superior, although participants would benefit from working faster even under low loads. Results from regressions and a simulation show that there is a mutually-reinforcing effect of system load and algorithm quality on not only whether but also *how* subjects use the algorithm, affecting service and waiting times. The interaction of the high load and superior algorithm lead subjects to serve customers faster than they would with the inferior algorithm or under low loads. We also test the effect of two interventions on participants' performance outcomes and show that these can improve service quality and speed, especially for workers facing low loads. We discuss the role of two mechanisms, learning by using and task reduction, that best explain our results, and the managerial implications for piloting new algorithms and implementing algorithms to promote service scale.

3 - Human-Adaptive Maintenance Optimization

Melvin Drent, Eindhoven University of Technology, Eindhoven, Netherlands, Rob Basten, Bas Van Oudenhoven, Collin Drent, Philippe Van de Calseyde

High-tech systems, infrastructure, and other capital assets require maintenance to remain operational. Sensors are installed to monitor the health of the asset, but these sensors are typically imperfect in that they may predict a breakdown of an asset when nothing is wrong, or they may predict that nothing is wrong, while a breakdown is imminent. We model this problem as a partially observable Markov decision process. Using sensor readings, we develop a belief of the underlying health condition of the asset. This belief can be used to warn a service engineer when it is time to inspect the asset, giving more accurate information on the underlying health condition. However, a service engineer may decide to ignore a warning if (s)he already received a few warnings, performed inspections, but did not find an issue. We perform an experiment to test how people react to warnings, i.e., to what extent they follow up the warnings and what determines their follow-up decision, e.g., the number of warnings that led to an inspection that was not necessary. We use our findings from the behavioral experiment to improve the model in that we optimize the warning signal, knowing how people may react to it. We perform a second experiment to see if our improved model leads to improved performance indeed.

SE41

Summit - 433

Experimental Design and A/B Tests in Practice

Invited Session

Applied Probability Society

Chair: NIAN SI, HKUST, Hong Kong, Hong Kong

Co-Chair: Zeyu Zheng, University of California, Berkeley, Berkeley, CA, United States

1 - Cluster-Randomized Designs for One-Sided Bipartite Experiments

Jennifer Brennan, Google Research, Kirkland, WA, United States, Jean Pouget-Abadie, David Holtz, Vahab Mirrokni

The conclusions of A/B tests may be biased when the outcome of one unit depends on the treatment status of other units, a problem known as interference. We study interference in the setting of one-sided bipartite experiments in which the experimental units do not interact directly. Instead, their interactions are mediated through their connections to interference units on the other side of a bipartite graph. This type of interference is common in marketplaces and two-sided platforms. We present a natural model for interference in one-sided bipartite experiments using the exposure mapping framework. Minimizing the bias of the difference-in-means estimator under our model results in a balanced partitioning clustering objective with appealing robustness properties. We explore the use of cluster-randomized designs in the setting of symbiosis bias in recommender systems, in which users (experimental units) experience interference due to an algorithm's recommendations of a shared corpus of content (interference units).

2 - Statistical Inference and A/B Testing for First-Price Pacing Equilibria

Luofeng Liao, Columbia University, NYC, NY, United States, Christian Kroer

We initiate the study of statistical inference and A/B testing for first-price pacing equilibria (FPPE). The FPPE model captures the dynamics resulting from large-scale first-price auction markets where buyers use pacing-based budget management. Such markets arise in the context of internet advertising, where budgets are prevalent.

We propose a statistical framework for the FPPE model, in which a limit FPPE with a continuum of items models the long-run steady-state behavior of the auction platform, and an observable FPPE consisting of a finite number of items provides the data to estimate primitives of the limit FPPE, such as revenue, Nash social welfare (a fair metric of efficiency), and other parameters of interest. We develop central limit theorems and asymptotically valid confidence intervals. Furthermore, we establish the asymptotic local minimax optimality of our estimators. We then show that the theory can be used for conducting statistically valid A/B testing on auction platforms. Numerical simulations verify our central limit theorems, and empirical coverage rates for our confidence intervals agree with our theory.

3 - Producer-Side Experiments Based on Counterfactual Interleaving Designs for Online Recommender Systems

Shan Ba, LinkedIn, Mountain View, CA, United States, Yan Wang

Recommender systems have become an integral part of online platforms, providing personalized recommendations for purchases, content consumption, and interpersonal connections. These systems consist of two sides: the producer side comprises product sellers, content creators, or service providers, etc., and the consumer side includes buyers, viewers, or customers, etc. To optimize online recommender systems, A/B tests serve as the golden standard for comparing different ranking models and evaluating their impact on both the consumers and producers. While consumer-side experiments is relatively straightforward to design and commonly employed to assess the impact of ranking changes on the behavior of consumers (buyers, viewers, etc.), designing producer-side experiments for an online recommender/ranking system is notably more intricate because producer items in the treatment and control groups need to be ranked by different models and then merged into a unified ranking to be presented to each consumer. Current design solutions in the literature are ad hoc and lacking rigorous guiding principles. In this paper, we examine limitations of these existing methods and propose the principle of consistency and principle of monotonicity for designing producer-side experiments of online recommender systems. Building upon these principles, we also present a systematic solution based on counterfactual interleaving designs to accurately measure the impacts of ranking changes on the producers (sellers, creators, etc.).

4 - Metric Decomposition in A/B Tests

Nathaniel Stevens, University of Waterloo, Waterloo, ON, Canada, Alex Deng, Luke Hagar, Tatiana Xifara, Amit Gandhi

More than a decade ago, CUPED (Controlled Experiments Utilizing Pre-Experiment Data) mainstreamed the idea of variance reduction leveraging pre-experiment covariates. Since its introduction, it has been implemented, extended, and modernized by major online experimentation platforms. Despite the wide adoption, it is known by practitioners that the variance reduction rate from CUPED utilizing pre-experimental data varies case by case and has a theoretical limit. In this talk, we propose a new direction for sensitivity improvement whereby a target metric of interest is decomposed into components with high signal-to-noise disparity. Inference in the context of this decomposition is developed using both frequentist and Bayesian theory. We provide three real world applications demonstrating different flavors of metric decomposition; these applications illustrate the gain in agility metric decomposition yields relative to an un-decomposed analysis.

SE42

Summit - 434

Navigating Distributions using Optimal Transport

Invited Session

Applied Probability Society

Chair: Bahar Taskesen, EPFL, Lausanne, Switzerland

1 - Empirical Martingale Projections via the Smoothed Adapted Wasserstein Distance

Johannes Wiesel, Carnegie Mellon University, Pittsburgh, PA, United States

Given a collection of multidimensional pairs $\{(X_i,Y_i):1\leq i\leq n\}$, we study the problem of projecting the associated suitably smoothed empirical measure onto the space of martingale couplings (i.e. distributions satisfying $\mathbb{E}[Y|X]=X$) using the adapted Wasserstein distance. We call the resulting distance the smoothed empirical martingale projection distance (SE-MPD), for which we obtain an explicit characterization. We also show that the space of martingale couplings remains invariant under the smoothing operation. We study the asymptotic limit of the SE-MPD, which converges at a parametric rate as the sample size increases if the pairs are either i.i.d. or satisfy appropriate mixing assumptions. Additional finite-sample results are also investigated. Using these results, we introduce a novel consistent martingale coupling hypothesis test, which we apply to test the existence of arbitrage opportunities in recently introduced neural network-based generative models for asset pricing calibration. (Joint work with J. Blanchet, M. Larsson, J. Park, E. Zhang, Z. Zhang)

2 -

Geometry of Constrained Mirror Flows: Insights and Applications in Diffusion Modeling

Ya-Ping Hsieh, ETH Zürich, Zürich, Switzerland

The emergence of diffusion models marks a significant milestone in data science, offering a versatile paradigm for tackling various tasks. Yet, despite their promise, developing efficient algorithmic solutions remains a formidable challenge. In addressing this, we propose an innovative approach that combines two fundamental mathematical frameworks: dynamical system theory and differential geometry. At the heart of our method lies a novel Riemannian geometric structure customized for probability measures, enabling the refinement of training schemes for diffusion models and thereby advancing their efficacy.

3 - Wasserstein Mirror Flows as Limit of the Sinkhorn Algorithm

Nabarun Deb, University of Chicago, Chicago, IL, United States, Young-Heon Kim, Soumik Pal, Geoffrey Schiebinger

We prove that the sequence of marginals obtained from the iterations of the Sinkhorn algorithm or the iterative proportional fitting procedure (IPFP) on joint densities, converges to an absolutely continuous curve on the 2-Wasserstein space, as the regularization parameter \varepsilon goes to zero and the number of iterations is scaled as 1/\varepsilon (and other technical assumptions). This limit, which we call the Sinkhorn flow, is an example of a Wasserstein mirror gradient flow, a concept we introduce here inspired by the well-known Euclidean mirror gradient flows. In the case of Sinkhorn, the gradient is that of the relative entropy functional with respect to one of the marginals and the mirror is half of the squared Wasserstein distance functional from the other marginal. Interestingly, the norm of the velocity field of this flow can be interpreted as the metric derivative with respect to the linearized optimal transport (LOT) distance. An equivalent description of this flow is provided by the parabolic Monge-Ampère PDE whose connection to the Sinkhorn algorithm was noticed by Berman (2020). We derive conditions for exponential convergence for this limiting flow. We also construct a Mckean-Vlasov diffusion whose marginal distributions follow the Sinkhorn flow.

4 - Distributionally Robust Control Using Optimal Transport

Bahar Taskesen, University of Chicago, Chicago, IL, United States, Dan Iancu, Cagil Kocyigit, Daniel Kuhn

Linear-Quadratic-Gaussian (LQG) control is a fundamental control paradigm that is studied in various fields such as engineering, computer science, economics, and neuroscience. It involves controlling a system with linear dynamics and imperfect observations, subject to additive noise, with the goal of minimizing a quadratic cost function for the state and control variables. In this work, we consider a generalization of the discrete-time, finite-horizon LQG problem, where the noise distributions are unknown and belong to Wasserstein ambiguity sets centered at nominal (Gaussian) distributions. The objective is to minimize a worst-case cost across all distributions in the ambiguity set, including non-Gaussian distributions. Despite the added complexity, we prove that a control policy that is linear in the observations is optimal for this problem, as in the classic LQG problem. We propose a numerical solution method that efficiently characterizes this optimal control policy. Our method uses the Frank-Wolfe algorithm to identify the least-favorable distributions within the Wasserstein ambiguity sets and computes the controller's optimal policy using Kalman filter estimation under these distributions.

SE43

Summit - 435

Learning and Control of Queueing Systems

Invited Session

Applied Probability Society

Chair: Xinyun Chen, Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A, China, People's Republic of

Co-Chair: Guiyu Hong, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A, China, People's Republic of

1 - Learning-Based Pricing and Matching for Two-Sided Queues

Zixian Yang, The University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Lei Ying

We consider a dynamic system with multiple types of customers and servers. Each type of waiting customer or server joins a separate queue, forming a bipartite graph with customer-side queues and server-side queues. The platform can match the servers and customers if their types are compatible. The matched pairs then leave the system. The platform will charge a customer a price according to their type when they arrive and will pay a server a price according to their type. The arrival rate of each queue is determined by the price according to some unknown demand or supply functions. Our goal is to design pricing and matching algorithms to maximize the profit of the platform with unknown demand and supply functions, while keeping queue lengths of both customers and servers below a predetermined threshold. This system can be used to model two-sided markets such as ride-sharing markets with passengers and drivers. The difficulties of the problem include simultaneous learning and decision making, and the tradeoff between maximizing profit and minimizing queue length. We use a longest-queue-first matching algorithm and propose a learning-based pricing algorithm, which combines gradient-free stochastic projected gradient ascent with bisection search. We prove that our proposed algorithm yields a sublinear regret $O(T^{5/6})$ and queue-length bound $O(T^{2/3})$, where T is the time horizon. We further establish a tradeoff between the regret bound and the queue-length bound: $O(T^{1-\gamma/4})$ versus $O(T^{\gamma})$ for $0<\gamma\leq 2/3$ (ignoring logarithmic factors).

2 - Online Learning and Optimization for Queues with Unknown Arrival Rate and Service Distribution

Guiyu Hong, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Xinyun Chen, Yunan Liu

We investigate an optimization problem in a queueing system where the service provider selects the optimal service fee p and service capacity μ to maximize the cumulative expected profit (the service revenue minus the capacity cost and delay penalty). The conventional predict-thenoptimize (PTO) approach takes two steps: first, it estimates the model parameters (e.g., arrival rate and service-time distribution) from data; second, it optimizes a model taking these parameters as input. A major drawback of PTO is that its solution accuracy can often be highly sensitive to the parameter estimation errors because PTO is unable to effectively account for how these errors (step 1) will impact the solution quality of the downstream optimization (step 2). To remedy this issue, we develop an online learning framework that automatically incorporates the aforementioned parameter estimation errors in the optimization process; it is an end-to-end approach that can learn the optimal solution without needing to set up the parameter estimation as a separate step as in PTO. Effectiveness of our online learning approach is substantiated by (i) theoretical results including the algorithm convergence and analysis of the regret ("cost" to pay over time for the algorithm to learn the optimal policy), and (ii) engineering confirmation via simulation experiments of a variety of representative examples. We also provide careful comparisons between PTO and our online learning method.

3 - Online Optimization for Network Resource Allocation

Yiqiang Zhao, Carleton University, Ottawa, ON, Canada

In this talk, we consider an on-line optimization problem in network resource allocation. In the network, there are N servers with resources which can be reserved at time t based on data at previous time slots to meet the unknown requests from the clients. Cost for reservation, for resource transfer, and also for violation occurs if the client request cannot be met. On-line algorithms based on randomized exponentially weights are proposed for minimizing the cost for both problems without and with a constrain, respectively. Performance of the algorithm is studied in terms of the regret.

This talk is based on joint work with Ahmed Sid-Ali, loannis Lambadaris, Gennady Shaikhet, and Amirhossein Asgharnia

4 - Dynamic Pricing for Make-to-Order Production Systems under Throughput Time Constraints: An effective computational method in the high-dimensional case

Chun Li, University of Chicago, Booth School of Business, Chicago, IL, United States, Baris Ata, NIAN SI

We consider a dynamic control problem in a make-to-order production setting, focusing on a multiclass, multi-server queueing system that serves multiple price-sensitive customer classes, each bound by specific throughput time constraints. To meet these constraints, the system manager can either dynamically adjust prices or reject orders when backlogs become excessive, thus incurring revenue losses. The objective is to minimize these discounted revenue losses through dynamic pricing, admission control, and scheduling decisions.

We approximate this problem by a multi-dimensional Brownian control problem in heavy traffic, and develop a neural network-based computational method tailored for high-dimensional problems. By interpreting the solution in the context of the original model, we propose an integrated dynamic pricing, admission control, and scheduling policy. The effectiveness of this policy is demonstrated through simulation.

SE44

Summit - 436

Bayesian Optimization

Invited Session

Simulation Society

Chair: Peter Frazier, Cornell / Uber, Ithaca, NY, 14850, United States

Co-Chair: Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

1 - Unexpected Improvements to Expected Improvement for Bayesian Optimization

Sebastian Ament, Meta, NYC, NY, United States, David Eriksson

Expected Improvement (EI) is arguably the most popular acquisition function in Bayesian optimization and has found countless successful applications, but its performance is often exceeded by that of more recent methods. Notably, EI and its variants, including for the parallel and multi-objective settings, are challenging to optimize because their acquisition values vanish numerically in many regions. This difficulty generally increases as the number of observations, dimensionality of the search space, or the number of constraints grow, resulting in performance that is inconsistent across the literature and most often sub-optimal.

Herein, we propose LogEI, a new family of acquisition functions whose members either have identical or approximately equal optima as their canonical counterparts, but are substantially easier to optimize numerically. We demonstrate that numerical pathologies manifest themselves in “classic” analytic EI, Expected Hypervolume Improvement (EHVI), as well as their constrained, noisy, and parallel variants, and propose corresponding reformulations that remedy these pathologies.

Our empirical results show that members of the LogEI family of acquisition functions substantially improve on the optimization performance of their canonical counterparts and surprisingly, are on par with or exceed the performance of recent state-of-the-art acquisition functions, highlighting the understated role of numerical optimization in the literature.

2 - Bayesian Optimization for Nonconvex Two-Stage Stochastic Optimization Problems

Jack Buckingham, University of Warwick, Coventry, United Kingdom, Juergen Branke, Ivo Couckuyt

Stochastic programming concerns optimization under uncertainty where average performance is the quantity of interest. In the first stage of a two-stage problem, here-and-now decisions must be made in the face of this uncertainty, while in the second stage, wait-and-see decisions are made after the uncertainty has been resolved. Typically, methods in stochastic programming assume that the objective is linear or convex, while Bayesian optimization allows to tackle non-convex problems that are expensive to evaluate. We present two knowledge-gradient based acquisition functions and demonstrate empirically that they outperform the naive approach of optimizing the two types of variables sequentially, both on synthetic and real-world examples.

3 - Pseudo-Bayesian Optimization

Haoxian Chen, Columbia University, New York, NY, United States

Abstract: Bayesian Optimization is a popular approach for optimizing expensive black-box functions. Its key idea is to use a surrogate model to approximate the objective and, importantly, quantify the associated uncertainty that allows a sequential search of query points that balance exploitation-exploration. Gaussian process (GP) has been a primary candidate for the surrogate model, thanks to its Bayesian-principled uncertainty quantification power and modeling flexibility. However, its challenges have also spurred an array of alternatives whose convergence proper- ties could be more opaque. Motivated by these, we study in this paper an axiomatic framework that elicits the minimal requirements to guarantee black-box optimization convergence that could apply beyond GP-related methods. Moreover, we leverage the design freedom in our framework, which we call Pseudo-Bayesian Optimization, to construct empirically superior algorithms. In particular, we show how using simple local regression, and a suitable “randomized prior” construction to quantify uncertainty, not only guarantees convergence but also consistently outperforms state-of-the-art benchmarks in examples ranging from high-dimensional synthetic experiments to realistic hyperparameter tuning and robotic applications.

4 - Gittins Indices as an Acquisition Function Design Principle in Bayesian Optimization

Alexander Terenin, Cornell University, Ithaca, NY, United States, Qian Xie, Raul Astudillo, Peter Frazier, Ziv Scully

Bayesian optimization is a popular class of black-box global optimization, which works by (a) building a probabilistic model for the objective function, and then (b) using the model to carefully decide where to evaluate the function gather data next by optimizing what is known as the acquisition function. Understanding how correctly design acquisition functions is a fundamental question in Bayesian optimization research, because choosing where to gather data next in the face of uncertainty leads to explore-exploit tradeoffs that must be balanced. The right balance can depend on the details of the precise setting at hand in non-obvious ways, and a thorough understanding of acquisition function design tends to only be available for problems with bandit feedback and other classical variants. In this talk, we show how Gittins index theory can provide new angles of attack for acquisition function design, especially in problems with explicit evaluation costs. Specifically, we map Bayesian optimization problems onto certain simplified analogues which can be solved exactly using Gittins index theory to derive an acquisition function for the setting at hand. In classical settings, the resulting acquisition functions are shown to perform strongly compared to baselines, especially once problem difficulty becomes large enough.

SE45

Summit - 437

Integrating Healthcare Research and Practice: Challenges and Opportunities

Panel Session

Health Applications Society

Co-Chair: Kanix Wang, University of Cincinnati, Cincinnati, OH, 45221, United States

1 - Moderator Panelist

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States

2 - Panelist

David Scheinker, Stanford University, Stanford, CA, United States

3 - Panelist

Pengyi Shi, Purdue University, West Lafayette, IN, United States

4 - Panelist

Amy Cohn, University of Michigan, Ann Arbor, MI, United States

5 - Panelist

Maria Mayorga, North Carolina State University, Raleigh, NC, United States

6 - Panelist

Joel Goh, NUS Business School, Singapore, Singapore

SE46

Summit - 438

Operations in Health Care and Public Sectors

Invited Session

Health Applications Society

Chair: Yichuan Ding, McGill University, Montreal, QC, H3A 1G5, Canada

1 - The Sooner, The Better? Optimal Vaccination Policy with Limited Vaccine Supply

Miao Bai, University of Connecticut, Storrs, CT, United States, George Chen, Cuihong Li

We study the optimal vaccination policy in an infectious disease outbreak, considering the limited supply of the vaccine and its imperfect efficacy. The inclusion of imperfect efficacy introduces an additional compartment to the celebrated Susceptible-Infectious-Recovered model, giving rise to an infinite horizon optimal control problem. To facilitate theoretical analysis, we propose a novel variable transformation that converts the problem into an equivalent form with linear dynamics. Leveraging this transformation, we derive a closed-form expression for the optimal vaccination policy under infinite administrative capacity and establish theoretical structures for the optimal policy under finite capacity. Our results suggest that delaying the vaccination process may be optimal, especially when the vaccine is less effective, the vaccine supply is more limited, and the disease is more infectious. The optimality of delay occurs because the individual-level benefit of vaccination with imperfect vaccines, in terms of the reduction of infection risk, is non-monotonic in time; hence maximizing the benefit of vaccination requires a strategic allocation of limited vaccine supply over time. Building on these theoretical findings, our numerical study demonstrates the significant benefit of delay in reducing the total number of infections compared to policies without delay. Our study contributes to the methodology of solving optimal control problems in the context of infectious disease outbreaks. Moreover, it highlights scenarios in which delaying the start of the vaccination process can be beneficial -- an aspect of vaccination policy design that has been overlooked in the literature but has important implications for practice.

2 - Stopping the Revolving Door: MDP-Based Decision Support for Community Corrections Placement

Xiaoquan Gao, Purdue University, West Lafayette, IN, United States, Pengyi Shi, Nan Kong

We study the incarceration-diversion decision problem to reduce jail overcrowding and recidivism rates by leveraging community corrections. We build a large-scale Markov decision process (MDP) model to balance the tradeoffs among congestion, recidivism, and violation. The salient features of the criminal justice setting, including deterministic service times and occupancy-dependent costs, present significant theoretical and algorithmic challenges. We propose a unified approach with system coupling and policy deviation bounding to compare value functions. We establish the superconvexity of the value function, which provides a theoretical basis for developing an efficient algorithm based on a separation of time scales. We showcase the effectiveness of our algorithm in solving real-world problems through a case study using data from our community partner.

3 - Optimizing Portability in Affordable Housing Allocation

Yifeng Cao, University of British Columbia, Vancouver, BC, Canada, Yichuan Ding, Daniel Granot, Hao Zhang, Weihua Zhang

This study investigates the allocation of affordable housing for low-income families, focusing on the decision-making process of housing authorities. We explore dynamic allocation models to optimize "portability", which refers to the easiness for an applicant's request being met. Findings indicate that myopic strategy and linear approximation have their respective domains of applicability. In particular, when the moving requests have stationary arrival rates, we prove that a myopic policy is asymptotically optimal; whereas the myopic policy can perform suboptimal under nonstatinary arrival rates. In the latter case, we propose a forward looking policy that achieves asymptotic optimality. Our findings offer insights for policymakers grappling with the complexities of affordable housing provision.

4 - Analyzing the Effects of Judicial Rotation on Criminal Sentencing: An Operations Management Perspective

Yuwei Zhou, University of Chicago, Chicago, IL, United States, Baris Ata, Rhys Hester, Lawrence Wein

A distinctive aspect of South Carolina's judicial system, which consisted of 50 judges presiding over 46 counties, was judicial rotation: although judges spent much of their time in their home circuit, they traveled to an average of 12 counties (typically holding court for a week in each county) and counties encountered an average of 13 different judges throughout the year. The empirical findings of Hester (2017) revealed that judicial rotation led to judge shopping, where defendants would strategically wait for a lenient judge before agreeing to a plea bargain. In this study, we construct a mathematical model that attempts to capture the effects of judge shopping. By simulating the mathematical model, we compute the mean and standard deviation of the sentence length as a function of three operational characteristics: the amount of judicial rotation, the allowable shopping time window for defendants, and the capacity utilization of the judicial system. We observe that both the mean and standard deviation of the defendant sentence length decreases in the amount of judicial rotation and the allowable shopping window for defendants, and increases in the capacity utilization. The average reduction is modest (≤ 10%), although a small proportion of defendants are impacted in a significant way. In a variant of the model adapted to an urban setting where all defendants have access to all judges, the mean and standard deviation of the sentence length decreases in the number of judges, even in the absence of intertemporal judge shopping.

SE47

Summit - 439

Pierskalla Best Paper Competition

Award Session

Health Applications Society

Chair: Ian Zhu, NUS Business School, Singapore, Singapore

Co-Chair: Rafid Mahmood, University of Ottawa + NVIDIA, Ottawa, ON, Canada

1 - Closer to Home: An Estimate-then-Optimize Approach to Improve Access to Healthcare Services

Jingyuan Hu, UCLA Anderson School of Management, Los Angeles, CA, United States, Fernanda Bravo, A Ghandi, Elisa Long

Geographic inequalities in access to essential health services extend beyond rural-urban differences to include socioeconomic, racial, and other disparities. Proximity to hospitals, clinics, and pharmacies varies greatly, complicating decisions about where to place these facilities to maximize access and utilization. Factors such as population health, individual preferences, and provider capacity all influence demand for services. This study introduces a novel estimate-then-optimize framework that combines structural demand estimation with an optimal facility location model. Unlike traditional approaches, this method uses aggregated data, such as market share, instead of relying on individual choices or outcomes, offering a more comprehensive understanding of health service needs.

2 - Policy Optimization for Personalized Interventions in Behavioral Health

Jackie Baek, New York University, New York City, NY, United States, Justin Boutilier, Vivek Farias, Jonas Oddur Jonasson, Erez Yoeli

We study the problem of optimizing personalized interventions for patients to maximize some long-term health outcome, in a setting where interventions are costly. We present a new algorithm we dub DecompPI, which consists of a prediction task from offline data. Theoretically, we show that under a natural set of assumptions on patient dynamics, DecompPI recovers at least 1/2 of the improvement between a naive baseline policy and the optimal policy. Through an empirical study on a mobile health platform for improving treatment adherence for tuberculosis, we find that DecompPI can provide the same efficacy as the status quo with approximately half the capacity of interventions.

3 - Large Language Models for Preventing Medication Direction Errors in Online Pharmacies

Cristobal Pais, University of California Berkeley, Berkeley, CA, United States, Jianfeng Liu, Robert Voigt, Vin Gupta, Elizabeth Wade, Mohsen Bayati

Errors in pharmacy medication directions can increase patient safety risks and operational inefficiencies. This study introduces MEDIC, a human-in-the-loop AI system that integrates domain knowledge with large language models (LLMs) to reduce these errors. MEDIC fine-tunes a first-generation LLM to extract core prescription components and assemble them into complete directions. In a retrospective study, MEDIC outperformed two LLM-based benchmarks. Prospectively, MEDIC reduced near-miss events —errors caught and corrected before reaching the patient— by 33% and enhanced overall operational efficiency.

4 - Disease Bundling or Specimen Bundling? Cost- and Capacity-Efficient Strategies for Multi-disease Testing with Genetic Assays

Hussein ElHajj, Santa Clara University, Santa Clara, CA, United States, Douglas Bish, Ebru Bish

We develop innovative models to optimize testing designs for infectious diseases that manifest with similar symptoms but often require different interventions. Our approach integrates multiplex assays (detecting multiple diseases from a single specimen) and pooled testing (combining specimens from multiple subjects). We construct the Pareto frontier for cost- and capacity-efficient solutions, derive structural properties and key insights on optimal designs, and develop efficient solution procedures for special cases. A case study on respiratory diseases, including COVID-19, demonstrates the advantage of our integrated, multi-disease testing design over existing methods.

SE48

Summit - 440

Cachon / Lariviere / Roels / Taylor Session

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Terry Taylor, U.C. Berkeley, Berkeley, CA, United States

1 - Payers and Waiters: Allocating Capacity in a Market with Line Sitting

Martin Lariviere, Northwestern University, Evanston, IL, United States, Achal Bassamboo

When capacity for an event is limited, consumers may queue for access. Alternatively, a consumer could hire someone to wait in their stead. Some thus compete via time and others via money. We examine such markets through a model in which consumers are subject to both time and monetary budget constraints. We show that the market outcome depends on whether consumers view time and money as substitutes or complements. When time and money are substitutes, line sitting does not change the allocation of who receives capacity. When they are complements, line sitting alters the market outcome with those earning low wages waiting for capacity while those earning high wages hire line sitters. Those with intermediate wages are squeezed out of the market.

2 - Why Do Managers Underdelegate? a Co-Productive Principal-Agent Model

Guillaume Roels, INSEAD, Fontainebleau, France, Vladimir Smirnov, Ilia Tsetlin, Andrew Wait

We consider a co-productive principal-agent (PA) model. Unlike traditional PA models, the principal can choose to work alone, i.e., to not hire the agent. We show that the problem of underdelegation stems from managers not partnering enough, but when they do, they tend to overdelegate the execution to the agent. Rather than controlling the agent, as the PA literature suggests, we propose to control the principal instead.

3 - The Enigma of Ticket Exchanges (and Other Reselling Markets)

Gerard Cachon, University of Pennsylvania, Philadelphia, PA, United States, Pnina Feldman

Abstract

The literature establishes (and practice confirms) that sellers can benefit from allowing consumers to purchase in advance of the date of actual consumption (e.g., concert tickets, sporting events, etc.). Because of this advance purchasing, consumers can find themselves either with a ticket that they no longer want, or without a ticket that they wish to have. In the past, scalpers would facilitate transactions among these consumers, for a fee. Sellers historically disliked those practices and actively worked to prevent them. In fact, we obtain a stark finding: an unfettered and efficient reselling market eliminates all of the benefits of advance selling, which justifies sellers' historic hostility to reselling. But now ticket exchanges are common, growing, and even embraced by the sellers. What changed? We present a theory that demonstrates reselling is actually beneficial for sellers under one crucial condition - the seller must have some control over the reselling process, thereby allowing the seller to earn something from each transaction through licensing fees to third-party sellers. The old-fashioned paper ticket did not give such control, but technology now enables electronic tickets, which do. In fact, a seller cannot earn more than what it receives from a properly designed reselling market (i.e., reselling is optimal for the seller), especially for popular events with limited capacity. Furthermore, speculators do not disrupt the market: i.e., the seller has no need for scalpers nor should fear them. In sum, our results explain why the seller's view towards reselling has shifted dramatically.

4 - What is the Impact of Labor Productivity on the Optimal Staffing Level?

Terry Taylor, U.C. Berkeley, Berkeley, CA, United States

No decision is more central to operations management than capacity, and no performance measure is more central to operations than productivity, i.e., output per unit time. What is the impact of this fundamental performance measure on this fundamental decision? This paper addresses this foundational question in the context of services: What is the impact of labor productivity on the optimal staffing level? To do so, the paper addresses one of the most basic questions that can be posed regarding a queueing system: What is the impact of the service rate μ on the optimal number of servers? The paper reveals that the answer exhibits a surprising, simple structure.

SE49

Summit - 441

Healthcare Analytics

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yangzi Jiang, The Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A, China, People's Republic of

1 - Adaptive Server Behavior to Schedule Deviations and Its Consequences: Evidence from Operating Rooms

Yiwen Jin, University of Calgary, Calgary, AB, Canada, Yichuan Ding, Steven Shechter, Jugpal Arneja

We investigate how clinical teams in operating rooms (ORs) adaptively respond to real-time deviations from their scheduled plans. We examine if clinical staff modify their service speed based on being ahead or behind schedule, and the implications of these adjustments on patient readmissions and reoperations.

Utilizing a unique dataset of surgery times, both actual and scheduled, we employ a dynamic panel model with the Arellano-Bond approach to identify adaptive behaviors and use an instrumental variable (IV) approach to address endogeneity issues, thus estimating the impact of surgical speed on quality.

We discover a new type of adaptive behavior in OR scheduling. Our findings show that surgical and cleaning teams increase their speed when they are behind schedule and decrease it when ahead. Specifically, surgical teams expedite surgeries by an average of 5.6% per standard deviation delay in the planned start, while taking 10.5% longer when ahead. Cleaning teams adjust their turnover times by accelerating 10.3% or slowing down 22.1% based on their scheduling status. These deviations are leveraged as IVs to demonstrate causally that speeding up surgeries beyond the planned schedule decreases surgical quality, indicated by higher rates of readmissions and reoperations within 30 days.

Our study highlights how surgical and cleaning teams' response to schedule deviations affects surgical quality. This insight can help hospital managers better predict end-of-shift times and plan surgeries to balance efficiency and quality, evidenced by a counterfactual analysis revealing a convex relationship between surgical outcomes and schedule discrepancies.

2 - Untangling the Effect of Physician Cross-Platform Exposure: Who Reap the Benefits?

Sijia Gao, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Tingting Song

A growing number of physicians are exploiting social media as a means of exposure to reach potential patients, deliver healthcare information, and offer guidelines for patients seeking healthcare services. This study investigates how physicians' content creation on social media influences patients' demands and satisfaction towards the physicians on online health platforms. Using the difference-in-differences approach, we uncover asymmetric influences of cross-platform spillovers for high- and low-awareness physicians. Specifically, low-awareness physicians do not enjoy the benefits (i.e., the increased volume of consultation orders) from content creation, but their ratings turn to decline due to attention distraction caused by cross-platform activities. Conversely, for high-awareness physicians, we find a positive cross-platform spillover effect for consultation orders without decreasing their ratings. A further test shows that despite the existence of attention distraction from cross-platform services for high-awareness physicians, the negative impact on feedbacks is offset by higher ratings from their cross-platform consumers. Our findings enrich previous literature on cross-platform spillover and provide insights into leveraging inter-platform impacts to reap benefits for physicians and online health communities.

3 - Efficiency Gains in the Health Sector Through Data-Driven Patient Scheduling: the Case of Paediatric Rehabilitation Service

Benjamin Ravenscroft, University of Waterloo, Waterloo, ON, Canada, Nasser Barjesteh, Hossein Abouee Mehrizi

Leveraging data from a large outpatient pediatric rehabilitation network, this study investigates two mechanisms affecting throughput in a multi-appointment healthcare setting: the number of appointments patients require and appointment attendance. First, we employ discrete time survival analysis methods to estimate the relationship between initial wait times and appointment modality and the number of appointments patients will require. We find that longer initial waiting times and mixing appointment modalities (i.e., some in-person, some virtual) is associated with patients requiring more appointments before discharge. However, these effects are modulated by patient severity. We also collected a large dataset of outpatient appointment booking data. This data was used to investigate the relationship between factors such as patients' distances from their clinic, appointment modalities, and the number of patients' known future bookings on the likelihoods of appointment attendance, no-show, and cancellation. Building upon these empirical results, we use a simulation study to assess the impacts of using priority-based waitlist queues on overall system throughput in a setting where wait times affect the number of appointments patients will require. Our findings yield important empirical insights in a multi-appointment service setting. Furthermore, we demonstrate the value of 'personalizing' scheduling systems in settings where client classes respond differently to waiting times and appointment modality.

4 - The Impact of Hospitals' Annual Award Display on Patient Consultation: Evidence from a Natural Experiment

Yangzi Jiang, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China, People's Republic of, Hao Wang, Jingqi Wang

Leveraging on an exogenous shock occurred on Haodf.com, a prominent medical consultation service platform in China, we explore how displaying the award sign influences the consultation volume in departments with different award status, including award-winning departments (AWD), non-awarded departments in hospitals with award (NAD), and departments in hospitals without any award (NAH). Haodf began displaying award signs (i.e., annual best hospital in various departmental categories) next to department names since November 2020. We also collected consultation data on the same set of hospitals and departments from Chunyu.com, another prominent online medical consultation platform in China. Using a difference-in-differences approach, we find that displaying award signs significantly increase the consultation volume in AWD, without reducing the consultation volume in NAD or NAH. In addition, the increase in consultations to AWD mainly comes from users with prior consultation experience with the focal department. Intriguingly, we also demonstrate that the financial barrier deters user migration from NAD to AWD. Our findings yield important managerial implications for expert service platforms by showing the effects of displaying the award sign on the user traffic distribution.

SE50

Summit - 442

Innovations in Health Care Delivery

Invited Session

MSOM: Healthcare

Chair: Kamalini Ramdas, London Business School, London, United Kingdom

Co-Chair: Nazli Sonmez, ESMT Berlin, Berlin, Germany

1 - Increasing the Follow-up Rate of Patients Requiring Emergency High-Consequence Treatment

Nazli Sonmez, ESMT Berlin, Berlin, Germany, Kamalini Ramdas

A low attendance rate for urgent, high-consequence treatment is prevalent in the developing world. In poorer countries, ignorance about the illnesses people have, allied to the financial costs of attending for treatment, are two of the main reasons why they fail to show up for medical appointments. Yet, for many diseases, patients may need surgical treatment soon after diagnosis to prevent severe disability. This project aims to identify ways to increase the attendance rate of patients who need emergency high-consequence treatment. Traditionally, in care systems across the world, information about why patients must undergo a particular treatment or care plan is communicated by emphasising medical aspects of the disease, typically in scientific terms. This research, in contrast, analyses the effect of providing emotionally engaging information about how a patient's life experience can be changed by their medical condition. The study runs a two-stage, randomised controlled trial at the world's largest eye hospital, the Aravind Eye Hospital in India. Although this research focuses on urgent high-consequence care contexts in the developing world, poorer patients in both developing and developed countries are less likely to show up for medical appointments for both urgent and non-urgent care. This study hopes to improve the chances of poorer populations receiving the care they need and to help reduce needless blindness. Furthermore, it hopes to inspire other healthcare providers globally to adopt the proposed methods.

2 - Improving Decision-Making in Resource Allocation: Evidence from Inpatient Admission Decisions

Harriet Jeon, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Song-Hee Kim, Hummy Song

Routing mechanisms are crucial for efficiently matching specialized resources to customer needs at the lowest cost. Improving allocation, then, is of first-order concern for many organizations. But when human decision-makers are tasked with allocation in a complex environment, it can be challenging to understand their decision processes and design systems that enhance their decision quality. Leveraging rich data derived from electronic medical records, we explore how, in the context of physicians making routing decisions of patients from the ED to inpatient medical units, decision-makers incorporate new information and adapt their decision-making processes.

3 - When Should Doctors and Patients Use Shared Decision-Making Under Bounded Rationality?

Feray Tuncalp, Bilkent University, Ankara, Turkey, Rouba Ibrahim, Song-Hee Kim, Jordan Tong

In recent years, shared decision-making (SDM), where doctors and patients jointly decide on appropriate treatment options, has drawn a lot of attention in the healthcare domain. SDM promises to improve patient utility, in part, through personalized care. Doctors inform patients about their personalized prognosis regarding risk and outcome predictions, and patients can incorporate their personalized lifestyle and risk preferences. To provide insights into when to personalize diagnoses and patient preferences, we study the decision-making process between two treatments using a stylized model where we account for both patients' and doctors' bounded rationality in decision-making, as well as doctors' miscalibration in their attempt to account for the random errors. Contrary to the common medical belief which advocates personalization of both medical diagnoses and patient preferences for all conditions, we find that there is no one-size-fits-all policy for the healthcare decision-making process.

4 - Impact of Empirical Research on Operations Management

Nalin Shani, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Maria Ibanez

Empirical research has contributed significantly to the field of operations management (OM). There has been a steady increase in empirical research being generated. Given this trend, an important question arises: how empirical research shapes the OM literature. We plan to investigate the impact empirical research has on recent publications.

SE51

Summit - 443

Commodities, Risk Management, Hedging, and Analytics

Invited Session

MSOM: iForm

Chair: Danko Turcic, The School of Business at the University of California, Riverside, Riverside, CA, United States

Co-Chair: Ye Liu, Syracuse University, Syracuse, United States

1 - Stochastic Capacity Investment in the Presence of Production Resource Uncertainty and Its Implications for Hedging

ONUR BOYABATLI, Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore, Guiyun Feng

Motivated by the recent incidents of production resource (component and financial resource) shortages, this paper studies a manufacturing firm's joint capacity investment and hedging decisions under demand and production resource uncertainties where the latter can be engineered by hedging. To this end, we consider a firm who produces and sells a single product in a single selling season to maximize its expected profit. We formulate a two-stage stochastic model. We conduct sensitivity analyses to examine the impact of production resource variability and its correlation with demand on the firm's optimal decisions and profitability as well as value of hedging. We identify correlation between demand and production resource uncertainty and the capacity investment cost as the key drivers of the optimal hedging decision. In particular, when the correlation is non-positive, the firm always fully hedges and production resource uncertainty is inconsequential for the firm. When the correlation is positive, full hedging is optimal only when capacity investment cost is sufficiently high. Otherwise, the firm chooses a partial hedging policy. Interestingly, the optimal partial hedge is chosen in such a way that there is no effect of production resource variability on the firm's profitability or capacity investment decision. We also find that the firm may optimally choose not to hedge at all, specifically, when the correlation is sufficiently high and the capacity investment cost is sufficiently low. Interestingly, we find that value of hedging decreases in correlation.

2 - Newsvendor&Rsquo;S Supplier Selection Problem with Correlated Supply and Demand Uncertainties

Liao Wang, The University of Hong Kong, Hong Kong, Hong Kong, Jin Yao, Sean Zhou, Hanyan Duan

We examine the supplier selection problem for a risk-averse newsvendor who faces stochastic demand and supply uncertainties. In addition to a single reliable supplier, we introduce multiple unreliable suppliers. Each unreliable supplier delivers the smaller of the ordered amount or its own stochastic capacity. Previous literature (Dada et al., A Newsvendor's Procurement Problem when Suppliers Are Unreliable, M&SOM, 2007, 9(1), 9-32) has established that with independent capacities and demand, a risk-neutral newsvendor will always prioritize cheaper suppliers (referred to as the "DPS's rule"). In this study, our objective is to maximize CVaR (Conditional Value at Risk) for the risk-averse newsvendor, while considering correlated demand and supply uncertainties. Our study reveals two key findings. First, we demonstrate that risk aversion alone does not influence the DPS's rule . Second, we establish general conditions under which the DPS's selection rule is reversed. Specifically, we identify a general sufficient condition, without assuming any specific relationship among the random variables, that leads to the selection of a more expensive supplier over cheaper ones. Our insight is that a positive relationship between demand and a supplier's capacity, combined with a negative relationship between this supplier's capacity and those of the other suppliers, can provide an advantage to the more expensive supplier and alleviate its cost disadvantage.

*This research is supported by the General Research Fund of Research Grants Council of HKSAR, with project code 17200521.

3 - Operational Resilience Anew

Andrea Roncoroni, ESSEC Business School, Cergy-Pontoise, France, Gianna Figà-Talamanca, Paolo Guiotto

Extreme events, such as pandemics or geopolitical crises, may lead to unexpected and sudden disruptions in the global supply chain. This reality may necessitate prompt and significant adjustments in physical operations, e.g., reshoring. Empirical evidence indicates that while many companies recognize this necessity, far fewer actually implement or are capable of doing so. In such contexts, academic research provides a range of operational tools to foster operational resilience. We propose a novel purely financial instrument to enable firms to enhance their operational resilience. Our approach is normative. We develop, analyze, and evaluate it using two integrated risk management (IRM) models. One is a stylized risk-averse newsvendor model featuring a decision maker contending with an unreliable supplier. We derive our IRM policy and empirically assess its effectiveness in bolstering operational resilience. The other model is a multinational capacity allocation model encompassing both reliable and unreliable suppliers. We calibrate it using historical gas market data and evaluate the resulting recommendations in light of the European gas supply chain disruption subsequent to the Russian attack on Ukraine. Our findings reveal that: (Operational contraction) A reduction in reliability prompts a decrease in optimal ordering/allocation; (Financially-driven resilience) IRM can mitigate this decrease compared to mere operational management; (All-weather feature) Traditional single-claim hedging offers uneven mitigation across reliability reductions compared to utilizing a combined custom hedge. In summary, our analysis indicates that the presence of an optimal IRM strategy mitigates the need for relocation, thereby offering managers a new avenue to strengthen operational resilience.

4 - Deep Hedging Strategies for a Risk Averse Hog Farm

Ye Liu, Syracuse University, Syracuse, NY, United States, Panos Kouvelis, Danko Turcic, Yunzhe Qiu

We investigate how a privately owned hog farm manages its inventory and uses hedging strategies when the farmer is concerned about risk. The farm faces risks from changing numbers of hogs, production costs, and the prices they can sell. The farm can use financial contracts to reduce these risks.

We find that a farmer who dynamically maximizes the mean-variance utility will keep a smaller inventory than a farmer who doesn't consider risk. However, they will keep a larger inventory than a farmer who doesn't have the option to hedge. We show that the decision to hedge revenue or costs depends on both operational and financial characteristics. Revenue hedging is more advantageous when there is high volatility in selling prices, a high correlation with factor markets, and large sales volumes. Conversely, cost hedging is more valuable in the presence of high volatility in operational costs, a high correlation with the fodder market, and large inventory sizes. We also find that a more risk-averse farm is likely to hedge costs, whereas a less risk-averse farm finds greater value in hedging revenue. Lastly, hedging is more valuable when the volatility of factor markets is lower. As the volatility of these markets increases, hedging loses its value since the risks from financial markets begin to propagate to the farm, contributing more variance to cash flow.

The hedging strategies are difficult to evaluate analytically because the farmer's problem is inherently dynamic, so we obtain the exact hedging policy through deep reinforcement learning.

SE52

Summit - 444

Ride-Hailing and Platform Operations

Invited Session

MSOM: Service Operations

Chair: Yuexing Li, Johns Hopkins Carey Business School, Baltimore, MD, 21202, United States

Co-Chair: Nur Sunar, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States

1 - Learning While Repositioning in on-Demand Vehicle Sharing Networks

Hansheng Jiang, University of Toronto, Rotman School of Management, Toronto, ON, Canada, Chunlin Sun, Shunan Jiang, Zuo-Jun Shen

We consider the vehicle repositioning problem for a one-way, on-demand vehicle sharing service with a fixed number of rental units distributed across the network. Due to uncertainty in both customer arrivals and vehicle returns, the service provider needs to periodically reposition the vehicles to match the supply with the demand while minimizing the total costs of repositioning labor and lost sales. The optimal repositioning policy under a general multi-location network is intractable without knowing the optimal value function. We define the best base-stock repositioning policy as a generalization of the popular inventory control policy to the vehicle repositioning problem, and we establish its asymptotic optimality in two different limiting regimes under general network structures. We provide reformulations to solve the offline problem of finding the best base-stock policy, and prove that the offline solution has favorable generalization properties. In the online setting, we develop an Online Gradient Repositioning algorithm utilizing only censored demand and, under a mild cost structure condition, we prove that it achieves an optimal regret, which matches the regret lower bound. Our online algorithm is designed by carefully decomposing the cumulative costs and formulating a linear programming problem whose dual solution serves as the gradient. Moreover, we provide and discuss three other online algorithms to elucidate the inherent challenges of learning while repositioning, and our analysis offers new insights into learning with censored data in networks. Numerical experiments illustrate the effectiveness of our proposed approach.

2 - Bundling OR Discounting? a Field Experiment for Healthy and Unhealthy Food

Nymisha Bandi, McGill University, Montreal, QC, Canada, Maxime Cohen, Saibal Ray

How can retailers decide between Bundling and Discounting as incentives to increase sales of healthy and unhealthy products? In this paper, we conducted a field experiment with a global convenience store chain to better understand how bundle and discount promotions influence healthy and unhealthy food choices. We considered six types of incentives: (i) discount a healthy product, (ii) discount an unhealthy product, (iii) bundle a healthy product with coffee, (iv) bundle an unhealthy product with coffee, (v) discount both healthy and unhealthy products, and (vi) choice bundle (offering either a unhealthy or a healthy snack with coffee). We conduct the field experiment in 7 different stores and observe the sales, revenue and profit impact for each of the above promotions on different category of products.

3 - Iot-Based Nudging for Energy Saving: More Can be Less for Organizations and Environment

Jacob Zeng, Gonzaga University, Spokane, WA, United States, Nur Sunar, Nil Karacaoglu, Serasu Duran

Excessive energy consumption results in substantial financial and environmental costs, yet the effectiveness of real-time feedback to correct overuse in commercial settings has not been investigated. In this paper, we collaborate with one of the largest Internet of Things (IoT) energy platforms in Europe to analyze a rich dataset on IoT-based nudging for energy saving.

4 - Greening Ride-Hailing: Impact on Environment and Consumers

Yuexing Li, Johns Hopkins Carey Business School, Baltimore, MD, United States, Bora Keskin, Nur Sunar

Ride-hailing services are an essential mode of transportation, providing convenience to millions worldwide. Many ride-hailing platforms have launched green initiatives to introduce and promote green vehicles. However, it remains a question of how adding green vehicles would impact the environment and consumers. To our knowledge, this paper is the first to analyze this question theoretically. We formulate a model where drivers decide on whether to serve local customers or reposition to serve customers in a different location. We consider two types of vehicles: regular (gas-fueled) and green. Some customers are environmentally sensitive and inclined to take rides in green vehicles, whereas others are inclined to take rides in regular vehicles. We find that, in contrast to common belief, adding more green vehicles to a ride-hailing platform may increase emissions. Our analysis also reveals that the environmentally optimal price margin for green vehicles can be positive or negative. Besides, we show that adding green vehicles may be a win-win for both customers for regular vehicles and those for green vehicles. Finally, we investigate several extensions, including the consideration of more locations, pricing for regular and green vehicles independently, pricing for profit maximization, and replacing regular vehicles while adding green ones. We verify that our main insights are robust to these scenarios. Our paper highlights that policymakers and ride-hailing platforms should make prudent decisions about introducing green vehicles to reduce emissions. Simply adding green vehicles without careful consideration may lead to unintended consequences.

5 - Dynamic Randomized FIFO Mechanism

Hongyao Ma, Columbia University, New York, NY, United States, Francisco Castro, Chiwei Yan

Ridesharing platforms match rider trips at airports to drivers who wait in a virtual queue. Under FIFO dispatching, the heterogeneity in trip earnings incentivizes drivers to cherrypick, increasing riders' waiting times and leading to a loss of efficiency and reliability. Castro et al. (2022) proposed a family of Randomized FIFO mechanisms that use waiting times to align incentives and reduce variabilities in drivers' total payoffs. The design of such mechanisms, however, requires full information on trip earnings and arrival rates for different destinations, which are challenging to estimate in practice. In this work, we introduce the Dynamic Randomized FIFO mechanism, which is very simple and does not rely on any information on market conditions.

We also develop a framework for computing approximate market equilibria in stochastic settings via best-response iterations. Extensive simulations demonstrate that our mechanism outperforms existing mechanisms guaranteed to achieve second-best outcomes in the fluid setting. Additionally, it shows robust performance when drivers exhibit heterogeneity in waiting costs or location preferences.

SE53

Summit - 445

Managing Supply Chain Risks with Emerging Considerations

Invited Session

MSOM: Supply Chain

Chair: Xin Geng, University of Miami, Coral Gables, FL, United States

Co-Chair: Guang Xiao, Hong Kong Polytechnic University, N/A

1 - Optimizing Procurement Tactics: Navigating Unreliable Suppliers and Quantity Constraint

Ning Ma, Arizona State University, Tempe, AZ, United States, Yimin Wang, Heng Zhang

As uncertainties such as yield, quality, and delivery duration continue to challenge supply chain functions, the imperative of supply base emerges. Observing the real-world trend of corporations streamlining their supplier bases, we seek to understand the best strategies for supplier selection and order allocation, especially when faced with uncertain delivery and a predefined upper limit on the number of suppliers selected from a broad pool. To this end, a multiple-supplier framework is proposed, reflecting suppliers as producers with varying reliabilities. We offer a quadratic approximation of the original problem and show the approximation's theoretical properties, enabling the proposal of a tractable algorithm and its fully polynomial-time approximation scheme (FPTAS).

2 - Adoption of Precision Agriculture Technologies in Farming

Shuhui Dong, Fudan University, Shanghai, China, People's Republic of, Lusheng Shao, Xiaole Wu

Precision agriculture (PA) is an emerging technology that intends to improve production efficiency and environmental sustainability by precisely allocating inputs to areas where crops require them most. Despite widespread encouragement from numerous countries, farmers' reluctance towards PA adoption persists. This hesitation primarily arises from a fundamental misunderstanding of how PA technologies can bolster their agricultural practices. In this study, we develop an analytical model to elucidate farmers' fertilizer application process and investigate the role of PA in their decision-making. We explore whether farmers should adopt PA when engaging in Cournot competition. Our findings reveal that PA enhances efficiency by reducing input losses and mitigating nutrient imbalance. However, it may intensify competition particularly when fertilizer costs are relatively high, thereby yielding mutual losses among farmers in the event of simultaneous PA adoption, even under complimentary adoption schemes. Furthermore, we find a mutually reinforcing relationship between adopting farmers, wherein the adoption of PA by one enhances the value for the other to adopt PA. Accordingly, we propose that governments with constrained financial resources stimulate widespread PA adoption by incentivizing initial adoption endeavors. We also explore the implications of PA across three dimensions: farmer welfare, environmental sustainability, and food security. Our analysis reveals that while PA does not consistently enhance farmer welfare or environmental outcomes, it emerges as a key driver of food security through its capability to augment crop output.

3 - Product Rollover and Timing Strategies in the Presence of Demand Uncertainty

Jian Li, Northeastern Illinois University, Chicago, Illinois, IL, United States, Panos Kouvelis

Within a stylized setting, we study the optimal timing and rollover strategy selection for the introduction of a new product. We examine two distinct timing strategies: the Static approach and the Dynamic approach. We consider two product rollover strategies: Solo-Rollover, and Dual-Rollover. Again, the selection of the rollover strategy can be made either statically, or dynamically over time.

Our focus is on formulating an optimization model that integrates the above-mentioned three layers of decisions in selecting strategy for product rollover and timing. We obtain a straightforward rule for determining whether the introduction of a new product is warranted in a given period and what is the optimal rollover strategy. Our decision approach relies on average demands, profit margins and the use of two straight lines segmenting the two-product demand space.

Within the rich modeling framework for the Dynamic Timing and Rollover Selection, we also study as special cases five other timing and rollover strategies that differ along the static/dynamic nature of timing of introduction and choice rollover approach.

4 - Sourcing Under Supply Uncertainty: Impact of International Commercial Terms

Jingwen Xu, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Yulan Wang, Guang Xiao, Lei Chen

Problem definition: In this study, we consider a three-tier decentralized global supply chain that consists of a domestic buyer, a reliable supplier located outside a free trade area, an unreliable supplier situated within a free trade area, and a logistics service provider that transports products from oversea suppliers to the domestic buyer. First, based on who shall bear the tariff and shipping cost, the buyer and the supplier can undertake one of the following three trade agreements: Ex Works, Delivered at Place, or Delivered Duty Paid. Second, depending on the basis of the dutiable value on which the tariff is calculated, Free on Board and Cost Insurance Freight are considered. Methodology: We propose a three-tier decentralized supply chain game and fully solve it to characterize each member's preference about the aforementioned trade agreements. Results: Among other results, we find that, as the tariff rate increases, counter to our intuition, the buyer is more willing to bear the tariff and freight rate. Interestingly, we also show that dual sourcing remains as the equilibrium for the buyer under FOB based tariff, whereas single sourcing from the unreliable supplier located in a free trade area can emerge when CIF is adopted and the tariff rate is relatively high. Managerial implications: Based on these findings, our paper proposes a short-term strategy associated with tariff adjustments and a long-term strategy with respect to trading term adjustments to help an import market keep business in its own free trade areas.

SE54

Summit - 446

Job Market Candidates' Session Focusing on Social Impact Operations at the Base of the Pyramid

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

Co-Chair: Tong Wang, London Business School, London, United Kingdom

1 - Ranking Quality and User Engagement on an Online B2B Platform

Rakesh Allu, Cornell University, Ithaca, NY, United States, Vishal Gaur

Online business-to-business (B2B) platforms are increasingly investing in data science teams to develop machine-learning algorithms for providing personalized rankings-based recommendations for their users. Consequently, there is growing need to develop ways to evaluate the performance of these algorithms. Two key questions faced by platforms in such evaluation are: (i) how to periodically measure the quality of rankings using real-time ranked transactions data generated by the users and, (ii) how to examine the effect of improving ranking quality on usage (i.e., the number of transactions) and browsing effort. These questions are challenging to address because of simultaneity between usage and effort and, censoring in ranked transactions data. Using detailed transaction-level data from a B2B platform in India for more than 61K sellers over one year, we propose methods to measure ranking quality, develop a position-level model of a user's decision to transact and scroll, and develop estimation approaches to overcome censoring. Our analysis reveals a position-wise dynamic that results in an increase in platform usage and a decrease users' browsing effort with improvement in ranking quality. We further show that nudging a user to scroll at least up to a predefined position can enable the platform to realize higher value from improving rankings quality. Our paper underscores the importance of operational improvements for platforms and highlights the need to pair ranking algorithms with appropriate behavioral prompts.

2 - Last-mile delivery of malaria prevention products in the DRC: An inventory management model under supply chain disruptions

ROBERT MONTGOMERY, International Rescue Committee, New York, NY, United States, Baris Ata, Amy Lehman

This paper studies inventory management policies to improve the last-mile delivery of healthcare products in remote settings prone to supply chain disruptions. It is motivated by the need for a reliable and cost-effective delivery system to resupply new mosquito-repellent products to combat malaria in the Lake Tanganyika region of The Democratic Republic of the Congo. The primary delivery methods used around Lake Tanganyika can become inoperable during periods of flooding, machinery breakdowns, and armed conflict. To ensure continuous coverage, our model allows for emergency delivery methods, with high variable costs, to be used during disruptions. In normal operating times, we use a large-capacity fixed-cost delivery method. We model the changes in the operating environment between normal and multiple disruption states using a Markov chain with disruption-specific parameters. We use an infinite-horizon discounted objective, which consists of ordering costs and a linear holding cost. We show that a novel periodic hybrid policy, which corresponds to an (s,S) policy for the normal environment state and a series of base-stock policies for the disrupted states, is optimal and provide the means to calculate the policy parameters. Additionally, we propose two easier to calculate heuristic policies based on widely used inventory management policies and show their relative effectiveness through simulation.

3 - Generative AI, Social Interaction, and Entrepreneurial Performance

Nicholas Otis, UC Berkeley, Berkeley, CA, United States

Access to high-quality information and advice is an important determinant of firm performance. As people increasingly turn to generative Artificial Intelligence (AI) for advice, it raises an important question: Is AI a complement or substitute for individuals' social networks? I explore this question through a seven-month-long pre-registered field experiment with small and medium enterprise (SME) entrepreneurs, which cross-randomizes access to an AI business assistant, a human business mentor, a combination of both, or neither. Among entrepreneurs assigned to receive access to AI alone, I find that AI crowds out use of entrepreneurs' business networks, leading to significant declines in business performance (profits and revenue). However, simultaneous access to both AI and a human mentor produces significant improvements in performance, which are considerably larger than the additive effect of the AI assistant and the human mentorship alone. These results suggest that individuals must leverage their networks in order to benefit from AI. However, access to AI alone crowds out those serendipitous conversations that help entrepreneurs discover new opportunities.

4 - Transmission Interaction Persistence (TIP): A Supply Chain and Epidemiological Model for Zoonotic Diseases Outbreaks

El Ghali Ahmed Zerhouni, Massachusetts Institute of Technology, Cambridge, MA, United States, Retsef Levi, Nicholas Renegar

Zoonotic viruses that jump from animals to humans have been the cause of major pandemics in the last several decades. Meanwhile farms could be a source of such diseases, most of the outbreaks have been linked to live animal markets in China. This suggests that markets potentially amplify zoonotic infections, a phenomenon that could not be explained by traditional epidemiologic models. This talk introduces a novel supply-chain motivated transmission, interaction, and persistence (TIP) stochastic epidemiology model that seeks to explain the mechanism of disease amplification in markets, with a case study on avian influenza. The analytical results demonstrate how major outbreaks could occur in markets, even when the infection rates among animals sourced from farms are very low, and the animals stay in the market for less than one day. This is the result of two important dynamics. The first of these dynamics concerns with the fact that markets serve as a consolidation point for agricultural inputs from many small farms. The other dynamics are based on the in-market operations that affect the environmental in-market infection evolution, particularly, persistent infection through two-sided interactions between animals and wastewater, surfaces, and feed. Thus, the study highlights the important role of supply chain structure and market operations in driving the spread of zoonotic viruses. Furthermore, there are direct managerial implications with respect to critical infection control strategies to reduce the risk of outbreaks, including aggressive and frequent sanitation of the market environment, as well as market design to reduce environmental infection.

SE55

Summit - 447

Community-engaged and Equity-forward Operations Research and Analytics I

Invited Session

Public Sector OR

Chair: Mubarak Iddrisu, University of Massachusetts Boston, Quincy, MA, 02169, United States

Co-Chair: Michael Johnson, University of Massachusetts Boston, Boston, MA, 02125, United States

1 - Operations Research and Analytics Applications for Community-Engaged Climate Response

Michael Johnson, University of Massachusetts Boston, Boston, MA, United States

Advancing climate justice agenda is a complex challenge facing policymakers and the society. Decision-support systems that are increasingly used to inform and aid policymaking process are rarely equipped to address equity concerns. Partly because decision-support systems are often built within disciplinary silos, do not recognize communities as experts, and prioritize efficiency over equity in model building exercises. In this position paper, we present a brief review of literature on six research fields that commonly engage with developing decision-support systems: i) Agent-based Modeling (ABM); ii) Artificial Intelligence; iii) Community-based Operations Research; iv) Decision-making under Deep Uncertainties; v) Green Information Systems; and vi) Systems Dynamics. We identify major challenges in each of these fields concerning climate justice agenda, with a specific focus to advance equity and transdisciplinarity with community-engagement. We also identify opportunities to build decision-support systems that can build on complementarity of each of these fields to overcome the limitations and suggest an integrative approach.

2 - Strategic Framework for Social Welfare Programs: Integrating Direct Cash and In-Kind Aid for Diverse Community Needs

Gulten Busra Karkili, University of Massachusetts-Amherst, Amherst, MA, United States, Senay Solak, Michael Johnson

The combination of expanding direct cash initiatives nationwide and existing in-kind programs targeting specific needs prompts the question of the optimal design of government policies to effectively assist those in need. Our study aims to develop a framework for designing benefit programs that cater to the diverse needs of targeted populations while balancing short-term urgencies with long-term wellbeing objectives. We propose an approach that integrates various forms of assistance, including direct cash transfers and in-kind aid, customized to specific recipient groups and their evolving circumstances. As we navigate resource constraints, our research explores methods for efficiently allocating funds towards long-term investments for unbanked and financially vulnerable communities, while simultaneously considering approaches to promote self-sufficiency. Through ongoing investigation, we seek to provide insights that will inform policymakers about the complexities of poverty alleviation strategies, with a focus on fostering sustainability and economic empowerment.

3 - Decision Modeling for Home-Matching Among Disadvantaged Residents

Karla Corres Luna, University of Massachusetts Boston, Boston, MA, United States

The urban housing crisis, driven by a lack of affordable homes and underused units, has not been effectively mitigated by government policies. In response, home-matching programs are emerging as promising solutions for low-income renters. In collaboration with Front Porch and using a community-engaged operations research approach, this study focuses on the housing challenges faced by marginalized groups such as the elderly, disabled, and low-income earners. The research goal is to apply problem-structuring methods to better understand the decision models and processes associated with home matchmaking using descriptive analysis of participant and organization data. This work yields insights regarding stakeholder values and identifies decision problems to improve housing outcomes for program participants. Moreover, the study seeks to formulate and solve stylized decision models based on field data. The problem structuring approach exercises and decision model solutions with simulated data can help determine the efficacy and social impacts of the Home Match business model and recommend improvements to Home Match's planning and operations. The findings will contribute to the literature by learning the complexities of the housing crisis and recommend effective strategies for promoting policies aimed at affordable housing solutions.

4 - Decision Modeling for Community-based Adaptation to Climate Change

Mubarak Iddrisu, University of Massachusetts Boston, Quincy, MA, United States

In response to increasing climate-driven hazards like floods, Massachusetts introduced the Municipal Vulnerability Preparedness (MVP) program to support local government efforts in climate adaptation planning and project prioritization. However, the MVP program does not meaningfully accommodate the experiences of disadvantaged residents with limited community involvement in flood adaptation regarding transit challenges. Thus, integrating the framework of community-based adaptation (CBA) with community-engaged operational research, we study flood-induced transit challenges faced by marginalized groups such as the elderly, disabled, and low-income earners. By exploring the gaps between expressed needs of disadvantaged residents and solution opportunities presented by MVP, we use problem structuring methods to identify a variety of decision problems in climate change adaptation, as defined by disadvantaged residents, with the belief that if solved, could improve their mobility needs. Moreover, there is a potential to apply community-engaged operational research principles to formulate and solve decision models that enable residents to generate alternative ways to cope with flood-induced transit challenges and achieve equitable mobility. The findings are anticipated to enhance the integration of equitable mobility and flood adaptation processes for marginalized communities in urban settings to adapt to climate change. The study contributes to the field of CBA in the urban context of a developed country by exploring its applicability and intersection with community-engaged operational research. Further, the community-engaged operational research framework bridges the gap between the various actors in CBA and serves as a decision tool for steering decision-making.

SE56

Summit - 448

MIF Early Career Award

Award Session

Minority Issues Forum

Chair: Gian-Gabriel Garcia, Georgia Institute of Technology, Atlanta, GA, United States

1 - Honorable Mention: Toward a Healthy, Fair, Resilient, and Connected World

Karmel Shehadeh, Lehigh University, Bethlehem, PA, United States

In this talk, we present an overview of our contributions toward a healthy, fair, resilient, and connected future. We focus on two main areas: (1) advancing stochastic and robust optimization methodologies with practical applications in healthcare, facility location, and transportation, and (2) developing fairness-promoting optimization frameworks. Additionally, we will highlight our service to the Operations Research and Management Science community, our dedicated efforts to advance diversity, equity, and inclusion, and our initiatives to promote STEM awareness among the younger generation.

2 - Winner: A Data-driven Journey to Health Equity

Karen Hicklin, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Join me as we journey along my career path where I use my academic training to analyze, model, and optimize systems and processes that impact the treatment and care of patients. I have dedicated my research to integrating operations research to model complex stochastic systems with a major focus on health equity for vulnerable and underserved populations. Thus far, my journey has led me to (1) develop strategies to evaluate public health programming and interventions, (2) use mathematical modeling for improved decision making, and (3) identify opportunities for intervention for improved health care quality and delivery. I will outline my contributions to maternal health, health disparities, and workforce development through my research and service interests.

SE57

Summit - Terrace Suite 1

Mathematical Modeling to Inform Decisions in Healthcare

Invited Session

Health Applications Society

Chair: Isabelle Rao, INSEAD, Fontainebleau, France

1 - Fast Stochastic Epidemic Simulations and An Adaptation of the Next Generation Matrix for a COVID-19 Epidemic Model of Social Distancing

Isabelle Rao, University of Toronto, Toronto, ON, Canada, Stephen Chick

Direct stochastic simulations of medium to large scale Markovian processes with population dynamics may have runtimes that are proportional to the population size, if they account for each state transition of each individual in the population. Several approaches to speed up such simulations have been proposed. We use a discrete-time, Euler-forward type approximation for state transition functions that simulates all transitions within a given time step in an effort to improve run times, at the expense of some (potentially correctable) bias. We illustrate this with a stylized model of COVID-19 social distancing interventions in the United Arab Emirates. We also adapt the next generation matrix method of Hill and Longini (2003) to a continuous time, discrete state model. The approach accelerates simulation run times from a linear scaling of run times in population size to a constant that depends on the number of possible state transitions.

2 - Cost-Effectiveness of Contingency Management for Methamphetamine use Disorder: a Model-Based Analysis

Gary Qian, Stanford University, Stanford, CA, United States, Yupeng Chen, Richard Rawson, Keith Humphreys, Margaret Brandeau

The escalating methamphetamine use disorder (MUD) in the United States represents a significant public health challenge. As of 2021, an estimated 1.6 million individuals aged 12 or older were diagnosed with MUD. Individuals with MUD can experience severe psychological and physical symptoms, including cognitive impairments and emotional dysregulation, and is often complicated by polydrug use, especially with opioids, leading to a sharp rise in overdose deaths.

Current treatment options for MUD are limited, as no pharmacotherapies have been approved by the FDA. Behavioral interventions, especially contingency management (CM), have emerged as effective alternatives. CM rewards individuals for maintaining drug-free behaviors, such as providing negative urine samples or attending therapy sessions. Despite strong evidence supporting its efficacy in reducing methamphetamine use and promoting health, CM faces significant implementation barriers. These include clinical challenges, such as a lack of comprehensive evidence and clinician involvement in program design, and political obstacles, such as concerns over the ethics of providing financial incentives in treatment.

We used a microsimulation model to evaluate the effectiveness and cost-effectiveness of CM in treating MUD.

3 - Incorporating Fairness In the Design and Evaluation of Capacity-Constrained Digital Health Interventions

Paul Dupenloup, Stanford University, Stanford, CA, United States, Margaret Brandeau, David Scheinker

The use of "just-in-time" digital health interventions is growing in clinical practice, however these are often deployed in pilot studies, rather than within a randomized clinical trial. The data collected in a pilot study can be used to optimize the intervention and a subsequent trial. However, in a capacity-constrained setting in which not all patients are eligible to receive a digital health intervention, there may be important tradeoffs between optimizing a clinical trial for predicted probability of success and health equity. Some patient populations could be left out of the study unintentionally if they are not expected to benefit significantly from the intervention, and may require an alternative or complementary intervention to achieve a greater treatment effect.

We leverage data from a remote patient-monitoring pilot study to design and compare interventions for a subsequent randomized clinical trial which balances predicted probability of success with health equity.

4 - Efficient Community Responder Crowdsourcing for TIME-Critical Medical Emergencies

Weiliang Liu, University of Chicago, Chicago, IL, United States, Timothy Chan, Zhisheng Ye

With the increasing prevalence of modern emergency response systems, crowdsourcing community responders near incident sites by sending alerts to them via mobile applications has become crucial for time-critical incidents like out-of-hospital cardiac arrest. An important decision under such circumstances is to determine the appropriate recipients of crowdsourcing alerts and the optimal timing for their dissemination. By far, there has been no systematic effort in developing analytical methods to understand the key trade-offs behind such decisions, which currently follow ad-hoc rules. This paper aims to help fill this gap, with the ultimate goal of providing analytically grounded responder alert strategies that achieve satisfactory response outcomes while minimizing excessive alerts.

SE58

Summit - Terrace Suite 2

Funding for Healthcare Research

Panel Session

Health Applications Society

Co-Chair: Saumya Sinha, University of Minnesota, Minneapolis, MN, United States

1 - Panelist

Oguzhan Alagoz, University of Wisconsin-Madison, Madison, WI, United States

2 - Panelist

Margaret Brandeau, Stanford University, Stanford, CA, United States

3 - Panelist

Mariel Lavieri, University of Michigan, Ann Arbor, MI, United States

4 - Panelist

Abdulaziz Ahmed, University of Alabama at Birmingham, Birmingham, AL, United States

SE59

Summit - Ballroom 1

APS Best Student Paper Competition

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Span-Based Optimal Sample Complexity for Weakly Communicating and General Average Reward MDPs

Matthew Zurek, University of Wisconsin-Madison, Highland Park, WI, United States

We study the sample complexity of learning an ε-optimal policy in an average-reward Markov decision process (MDP) under a generative model. For weakly communicating MDPs, we establish the complexity bound SAH/ε^2 (ignoring constants and log factors), where H is the span of the bias function of the optimal policy and SA is the cardinality of the state-action space. Our result is the first that is minimax optimal (up to log factors) in all parameters S,A,H, and ε, improving on existing work that either assumes uniformly bounded mixing times for all policies or has suboptimal dependence on the parameters. We also initiate the study of sample complexity in general (multichain) average-reward MDPs. We argue a new transient time parameter B is necessary, establish an $SA(B+H)/\varepsilon^2$ complexity bound, and prove a matching (up to log factors) minimax lower bound. Both results are based on reducing the average-reward MDP to a discounted MDP, which requires new ideas in the general setting. To optimally analyze this reduction, we develop improved bounds for γ-discounted MDPs, showing that $SAHE^2/\varepsilon^2$ and $SA(B+H)E^2/\varepsilon^2$ samples suffice to learn ε-optimal policies in weakly communicating and in general MDPs, respectively, where $E =1/1-\gamma$ is the effective horizon. Both these results circumvent the well-known minimax lower bound of $SA E^3/\varepsilon^2$ for γ-discounted MDPs, and establish a quadratic rather than cubic horizon dependence for a fixed MDP instance.

2 - Random Graph Matching with Multiple Graphs

Taha Ameen, University of Illinois Urbana-Champaign, Urbana, IL, United States

This talk concerns fundamental limits for graph matching, i.e., recovering the underlying hidden correspondence among the nodes of multiple correlated random graphs. We identify a necessary condition for any algorithm to correctly match all nodes across all graphs and propose two algorithms for which the same condition is also sufficient. The first algorithm employs global information to simultaneously match all the graphs, whereas the second algorithm first partially matches the graphs pairwise and then combines the partial matchings by transitivity. We show that the second algorithm is an efficient and optimal bridge between 2-ary and m-ary graph matching, and that both algorithms work down to the information theoretic threshold. Our analysis reveals a scenario where exact matching between two graphs alone is impossible but leveraging more than two graphs allows exact matching among all the graphs. Along the way, independent results about the k-core of Erdős-Rényi graphs are derived.

3 - Strongly Tail-Optimal Scheduling in the Light-Tailed M/G/1

George Yu, Cornell University, Ithaca, NY, United States

We study the problem of scheduling jobs in a queueing system, specifically an M/G/1 with light-tailed job sizes, to asymptotically optimize the response time tail. This means scheduling to make P[T>t], the chance a job's response time exceeds t, decay as quickly as possible in the limit. For some time, the best known policy was First-Come First-Served (FCFS), which has an asymptotically exponential tail: P[T>t]~Ce−γt. FCFS achieves the optimal *decay rate* γ, but its *tail constant* C is suboptimal. Only recently have policies that improve upon FCFS's tail constant been discovered. But it is unknown what the optimal tail constant is, let alone what policy might achieve it. In this paper, we derive a closed-form expression for the optimal tail constant C, and we introduce *γ-Boost*, a new policy that achieves this optimal tail constant. Roughly speaking, γ-Boost operates similarly to FCFS, but it pretends that small jobs arrive earlier than their true arrival times. This significantly reduces the response time of small jobs without unduly delaying large jobs, improving upon FCFS's tail constant by up to 50% with only moderate job size variability, with even larger improvements for higher variability. While these results are for systems with full job size information, we also introduce and analyze a version of γ-Boost that works in settings with partial job size information, showing it too achieves significant gains over FCFS. Finally, we show via simulation that γ-Boost has excellent practical performance.

4 - A Hierarchical Approach to Robust Stability of Multiclass Queueing Networks

Feiyang Zhao, ExxonMobil, Houston, TX, United States

The focus of this research is on the robust stability of stochastic processing networks, under a wide class of control policies. A framework for robust stability is developed, in which the resources are given freedom to decide their own priority policy, under some general constraints. We offer a new approach for determining sufficient conditions to robust stability, which builds on, and makes connections to, suitable robust optimization problems, with the collection of priority policies as the uncertainty set. We also show how robust stability of a family of policies is inherited from the stability of some special policies, i.e., static-priority policies.

SE60

Summit - Ballroom 2

ML and Optimization Driven Fulfillment Operations Management

Invited Session

The Practice Section of INFORMS

Co-Chair: PREM KUMAR VISWANATHAN, Amazon, Seattle

1 - Warehouse Order Picking and Container Fill Optimization in Amazon Grocery Operations

Prem Kumar Viswanathan, Amazon, Seattle, WA, United States

This paper focuses on the problem of picking orders for Amazon Grocery with the twin objectives of lowering fulfillment costs and transportation costs. The fulfillment cost optimization is achieved by lowering travel distance between two picked items, while the transportation cost optimization is achieved by improving the fill rate on the containers used to haul these picked items, thereby increasing the cubic volume moved on trucks. We developed mathematical models utilizing the decomposition techniques and multi-stage solution approaches as well as variations of three-dimensional bin packing heuristics to solve the models. The integrated warehouse picking and container fill optimization problem has not been studied in the literature and we believe this seminal work will help pursue research in this area. Both mathematical models and heuristic approaches have been implemented and tested on over 700 problem instances. Mathematical models outperform the heuristic approaches for both fill-rate and travel time between picks.

2 - Visual-Assisted Approach to Optimizing Picking Operations in Fulfillment Centers

Akshay Kurapaty, Amazon, Bellevue, WA, United States, Abhisek Mukhopadhyay, SMA Bin Al Islam, Kay Zheng, Elcin Cetinkaya

In the dynamic landscape of e-commerce warehousing, optimizing picking processes is crucial for enhancing operational efficiency. Our research demonstrates that items with better visibility and accessibility significantly contribute to faster pick rates. Items that lack visibility or are obstructed by other items within a bin prove to be less suitable for fulfilling pick requests efficiently. We introduce a novel learning-based approach to assess pickability of an item. Pickability refers to the measure of how suitable an item is for being picked from a specific bin or shelf, within a given operational context. A high pickability score implies that an item is well-positioned, visible, and easily retrievable based on the state of a bin. This score encapsulates various factors, including item accessibility and visibility within bins, tailored to the specific operational context of the warehouse. By injecting pickability scores into the pick scheduling algorithm, we effectively guide the associates to bins containing items with shorter identification times. Consequently, this proactive approach leads to tangible improvements in overall pick time.

3 - Ergonomics Aware Stow Assistance

Abhisek Mukhopadhyay, Amazon, BELLEVUE, WA, United States, Andrew Johnson

In the Stow Process, associates at Amazon Robotics Sortable Fulfillment Centers are tasked with stowing individual items into bins in storage pods, from where pickers later retrieve them to fulfill customer orders. The pods are notably taller than an average human. Extreme locations within these pods pose ergonomic challenges during stowing and picking activities, with frequent bending and reaching movements correlating with an increased risk of Musculo-Skeletal Disorder (MSD) related injuries in associates. This paper presents Machine Learning and Simulation Model based solution to control MSD risk factors by guiding associates to stow fast-moving inventory in more ergonomic locations, while stowing slower-moving inventory to less ergonomic areas. Through experiments conducted in 5 fulfillment centers across North America, a 20-22% reduction in stows at the least ergonomic locations was observed, resulting in a statistically significant decrease in MSD risk. This study demonstrates that redistributing inventory over time enhances transaction frequency in ergonomically advantageous locations, contributing to a safer working environment for stow associates

4 - Error-Based Forecast Refinement: Bridging Past Mistakes to Enhance Future Predictions

Chinmoy Mohapatra, Amazon, Lake Stevens, WA, United States, Rohit Malshe, Abhilasha Katariya

Accurate forecasting is crucial for informed decision-making, yet discrepancies between forecasts and actual outcomes can lead to suboptimal plans and erode trust in predictive models. This paper presents a novel bridging methodology to analyze errors in any forecasting model and attribute them to specific input inaccuracies or model deficiencies using the Taylor approximation. We demonstrate its value for bridging last mile efficiency forecasts for Amazon Logistics. This scalable and automated error-attribution process has enhanced future planning and efficiency predictions by enabling targeted improvements. Since deployment in 2021, it has significantly improved Amazon's last mile planning process by identifying the best areas of opportunity to work on and has reduced the planning gap from by almost half through various science and data engineering efforts. This research aims to improve model explainability, enhance trust in predictive systems, and enable data-driven model refinement.

SE61

Summit - Ballroom 3

Educating in the Age of AI: Future-proofing OR Practitioners

Panel Session

Committee's Choice

Co-Chair: Ranganath Nuggehalli, UPS, Timonium

Co-Chair: Anne Robinson, Robinson Insights, Ottawa, ON, Canada

1 - Moderator Panelist

Anne Robinson, na, Ottawa, ON, Canada

2 - Moderator Panelist

3 - Panelist

Wenjun Zhou, University of Tennessee Knoxville, Knoxville, TN, United States

4 - Panelist

Brad Warner, United States Air Force Academy, Air Force Academy , CO, United States

5 - Panelist

Michael Fry, University of Cincinnati, Cincinnati, OH, United States

6 - Panelist

Maria Vlasiou, University of Twente, Eindhoven, Netherlands

7 - Panelist

Harish Krishnan, University of British Columbia, Vancouver, BC, Canada

SE62

Summit - Signature Room

Machine Learning Methods for Large Population Games

Invited Session

TutORial

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Machine Learning Methods for Large Population Games

Gokce Dayanikli, University of Illinois Urbana-Champaign, Champaign, IL, United States, Mathieu Lauriere

In this tutorial, we provide an introduction to machine learning methods for finding Nash equilibria in games with large number of agents. These types of problems are important for the operations research community because of their applicability to real life situations such as control of epidemics, optimal decisions in financial markets, elec tricity grid management, or traffic control for self-driving cars. We start the tutorial by introducing stochastic optimal control problems for a single agent, in discrete time and in continuous time. Then, we present the framework of dynamic games with finite number of agents. To tackle games with a very large number of agents, we discuss the paradigm of mean field games, which provides an efficient way to compute approxi mate Nash equilibria. Based on this approach, we discuss machine learning algorithms for such problems. First in the context of discrete time games, we introduce fixed point based methods and related methods based on reinforcement learning. Second, we discuss machine learning methods that are specific to continuous time problems, by building on optimality conditions phrased in terms of stochastic or partial differential equations. Several examples and numerical illustrations are provided along the way.

SE63

Regency - 601

Secondary Effects of the (Generative) AI Revolution

Invited Session

Information Systems

Chair: Martin Quinn, Erasmus University, Rotterdam, N/A, Netherlands

Co-Chair: Dainis Zegners, Rotterdam School of Management, Erasmus University, No mail please, 3068KE, Netherlands

Co-Chair: Martin Quinn, Rotterdam, Netherlands

1 - Does AI Technology Deployment Benefit the Owner of the Technology? Impact of GitHub Copilot Release on Microsoft

Poonacha Medappa, Tilburg University, Tilburg, Netherlands, Murat Tunc, Feiyang (Amber) Xu

The deployment of new AI technologies usually occurs in two phases: the technical preview, followed by the public release. These two phases differ in terms of access to AI technology. During the technical preview, access is limited to the organization that owns the technology, while it becomes widely available in the public release phase. This paper investigates the impact of these two deployment phases on the organization that owns the technology. Using a panel dataset spanning from 2020 to 2023, we examine the two-phase deployment of GitHub Copilot (as a technical preview in June 2021 and as a public release in June 2022) and how it impacts Microsoft's (the owner of GitHub) open-source software projects. Results show that during the technical preview, Microsoft projects experienced significant increases in development activity (commits, code additions, file changes) compared to the control group consisting of Google, IBM, and Meta. However, this advantage diminished after the public release. While the technical preview does not significantly impact project maintenance (issues reported, resolution time), the public release led to significant increases in both measures for Microsoft. The findings suggest that the technical preview of an AI technology benefits the owner organization's development productivity, but these gains are reduced, and maintenance activities emerge following the public release. Our study highlights the importance for organizations to develop curated deployment strategies for their AI technologies to maximize benefits and mitigate potential drawbacks associated with different phases.

2 - Strategic Behavior and AI Training Data

Jeremie Haese, University of Lausanne, Lausanne, Switzerland, Christian Peukert, Florian Abeillon, Franziska Kaiser, Alexander Staub

Human-created works represent critical data inputs to artificial intelligence (AI). Strategic behavior can play a major role for AI training datasets, be it in limiting access to existing works or in deciding which types of new works to create or whether to create any at all. We examine creators' behavioral change when their works become training data for AI. Specifically, we focus on contributors on Unsplash, a popular stock image platform with about 6 million high-quality photos and illustrations. In the summer of 2020, Unsplash launched an AI research program by releasing a dataset of 25,000 images for commercial use. We study contributors' reactions, comparing contributors whose works were included in this dataset to contributors whose works were not. Our results suggest that treated contributors left the platform at a higher-than-usual rate and substantially slowed down the rate of new uploads. Professional and more successful photographers had a stronger reaction than amateurs and less successful photographers. We also show that affected users changed the variety and novelty of contributions to the platform, which can potentially lead to lower-quality AI outputs in the long run. Taken together, our findings highlight the trade-off between interests of rightsholders and promoting innovation at the technological frontier. We discuss implications for copyright and AI policy.

3 - The Subtle Persuasion of Generative AI on Opinion Formation

Sterling Ceci-Williams, Cornell Tech, New York City, NY, United States, Maurice Jakesch, Advait Bhat, Kowe Kadoma, Lior Zalmanson, Mor Naaman

The talk presents an analysis of studies (n=3,024) that show the influence of AI-generated autocomplete suggestions on shaping views regarding various societal issues. Despite being exposed to biased AI suggestions, participants remained oblivious to how these biases shifted their expressed opinions. This phenomenon persisted even when individuals were forewarned about potential biases, underscoring the challenge of protecting human autonomy from the subtle yet powerful sway of AI.

These studies highlight the complex interplay between AI and human decision-making, showing a consistent inclination to defer to AI in cases of complex opinion formation, even against the backdrop of manifestly biased AI suggestions. The insights gathered call for a critical dialogue on the imperative to cultivate strategies that reinforce human autonomy, urging a deliberate and mindful integration of AI that safeguards independent human judgment in the face of expanding AI influence.

SE64

Regency - 602

Artificial Intelligence and Digital Platforms

Invited Session

Social Media Analytics

Chair: Xue Tan, Southern Methodist University, Dallas, 75252, United States

1 - Patched Vulnerability Disclosure and Hacker Participation in Bug Bounty Programs

Ali Ahmed, Louisiana State University, Baton Rouge, LA, United States, Brian Lee, Amit Deokar

In cybersecurity, bug bounty programs (BBP) have emerged as an effective method of finding hidden vulnerabilities in a system. In this method, ethical hackers discover and report hidden vulnerabilities to the organizations. In return, these hackers receive rewards for their valid discoveries. The emergence of bug bounty programs has enabled the practice of Patched Vulnerability Disclosure (PVD), where firms disclose vulnerability information after they have been patched. Currently, there is little understanding of how PVD affects ethical hackers' participation in BBPs. We collected a dataset from a leading bug bounty platform and conducted empirical analyses to analyze the effect of PVD on hacker participation. Leveraging multiple specifications, including an instrumental variable approach, difference-in-differences, and data collected from third-party websites, our analysis reveals that as firms disclose more patched vulnerabilities, hacker participation in their programs decreases. To further explore the mechanism, we found that disclosing valid vulnerability reports discourages hacker participation, whereas disclosing invalid reports leads to more hacker participation. We found that PVD influences uncertainty in the bug bounty process, which can simultaneously impact the risk and ambiguity in hacking. Valid disclosures can escalate hackers' risk and discourage participation, whereas disclosing invalid submissions can reduce the ambiguity in the bug bounty process and increase the hacker's participation in BBP. We also found that reduced ambiguity is most salient for hackers with experience working with other firms on the platform. Overall, our study makes theoretical contributions to the literature on vulnerability disclosure, ethical hacking, and the management of bug bounty programs.

2 - Chaos is a Ladder: Artificial Intelligence and Employee Satisfaction

Yingying Chen, University of Washington, Seattle, WA, United States, Wenxin Huang, Tingting Song

This paper addresses the unresolved question of how artificial intelligence (AI) impacts human well-being. Employing textual analysis techniques, we develop a novel measure of AI exposure for jobs and create two dummy variables to distinguish between occupations likely to be complemented or substituted by AI technology, utilizing ChatGPT. We investigate the influence of AI exposure on employee satisfaction, finding that increased exposure correlates with higher levels of employee satisfaction, with self-esteem and social relatedness acting as partial mediators. Conversely, individuals in roles more susceptible to AI replacement exhibit signs of depression. Leveraging AI regulatory policy as a quasi-natural experiment, we employ various difference-in-differences methods to establish a causal relationship between AI and employee well-being, both within and beyond organizational boundaries. In summary, our findings suggest that the rise of AI makes most people better off, but those who may be replaced by AI technology are concerned about the future, which exacerbates the gap in employees' subjective well-being.

3 - AI-Powered Digital Streamers for Online Retail: Empirical Evidence and Design Strategies Through Experiments

Lei Wang, Pennsylvania State University, University Park, PA, United States, Yahui Liu, Yanwen Wang

Artificial intelligence (AI) advancements have sparked a rise in AI-powered digital streamers for live streaming, significantly transforming online retail. Despite their potential for cost-effective, customizable, and reliable product promotion, it remains unclear whether they can rival real human streamers in sales performance and what design features could boost their effectiveness. This study aims to address these questions by examining the impact of digital streamers on product sales and identifying design strategies to improve their performance. In the first study using a quasi-experiment with a leading fashion retailer, we discovered that the plain villa digital streamers had no significant increase in sales as compared to a baseline scenario without streaming. To understand how to enhance digital streamers' sales effectiveness, our second study, a randomized field experiment, examined various design strategies for digital streamers. Our findings indicate that scripts trained from expert human hosts do not significantly increase sales compared to plain villa digital streamers. In contrast, improvements in the streamers' appearance and voice lead to sales increases of 12% and 17%, respectively. Additionally, incorporating lottery boosts sales by 17%, and enhanced real-time Q&A interaction is the most effective strategy, with a 26% increase in sales. This performance is equivalent to the 31% sales increase achieved by real human counterparts. This study contributes to the literature by enhancing our knowledge of how design features affect consumer behavior and underscores the benefits of human-AI interaction in e-commerce. We also offer managerial insights for online retailers and digital streamer designers on optimal design strategies.

4 - Driver on Deck: Impact of Order Assignment in On-Demand Delivery Platforms

Caitlin Cunningham, University of Washington, Seattle, WA, United States, Leela Nageswaran

On-demand delivery platforms feature unique traits such as the possibility of stacking orders and communicating progress to consumers. Using data from a food delivery platform, we empirically study the impact of these traits on consumers. Our results provide guidance to platforms regarding the implications of order stacking on the order lifecycle.

SE65

Regency - 603

NSF Proposal Writing Workshop

Invited Session

NSF

Chair: Georgia-Ann Klutke, National Science Foundation, Alexandria, VA, United States

1 - NSF Proposal Writing Workshop

Georgia-Ann Klutke, National Science Foundation, Alexandria, VA, United States

This session will provide guidelines and tips for proposal preparation for funding consideration by the National Science Foundation. Intended primarily for early-stage investigators, the presentation will review proposal preparation and submission, NSF’s Intellectual Merit and Broader Impacts criteria, proposal processing and review, and post-decision follow up. A Question-and-Answer session will follow the presentation.

2 - NSF Proposal Writing Workshop

Reha Uzsoy, National Science Foundation, Alexandria, VA, United States

This session will provide guidelines and tips for proposal preparation for funding consideration by the National Science Foundation. Intended primarily for early-stage investigators, the presentation will review proposal preparation and submission, NSF’s Intellectual Merit and Broader Impacts criteria, proposal processing and review, and post-decision follow up. A Question-and-Answer session will follow the presentation.

SE66

Regency - 604

Advancements in Time Series Analysis

Contributed Session

Contributed

Chair: Ming Luo

1 - Machine Learning for Time Series Analysis and Forecasting

Ming Luo, northeastern university , Boston, MA, United States

Time series data are prevalent and essential in decision-making. Author applies machine learning techniques to analyze time series data for classification, clustering, and forecasting. First, a new distance measure, value-added, is proposed in time series classification and clustering. Further, the author develops a novel framework in which decisions such as the number of clusters and prediction based on value-added are made using different techniques.

Moreover, forecasting in scale is a particular issue in business forecasting. A new approach reconciliation with neural network (R-NN) is proposed for hierarchical forecasting with the base forecasting from previous time series classification and clustering process, considering the non-linear relationship among time series for forecasting. Conventional techniques have yet to incorporate the relationship between series while making individual predictions and tend to lose information. In addition, computational costs can be exceptionally high due to model perplexity when using methods such as optimal reconciliation. The R-NN is straightforward to implement and fast to train without losing information.

The methods and procedures developed here can be applied in various business settings. Especially for companies with limited labor resources to make predictions on plenty of products, this approach can still generate a good prognosis for all the series. Likewise, the reconciliation with neural network (R-NN) can be used to forecast many time series simultaneously consistently. Then a coherent hierarchy of forecasts can assist subsequent supply chain decisions such as production and deployment.

2 - Unsupervised Anomaly Detection with Multivariate Time Series: Leveraging Mixture of Experts and Time Series Decomposition

Soomin Lee, Chungnam National University, Yuseong-gu, Daejeon, Korea, Republic of, Dongil Kim

Multivariate time series(MTS) is crucial in industries utilizing multiple sensors. Anomaly detection with MTS is one of the most important tasks. Anomaly detection typically employs unsupervised methods due to the absence of labeled data. A common approach is to train models solely on normal data and identify anomalies based on the degree of reconstruction. In this study, we propose an unsupervised anomaly detection method with MTS. Specifically, the proposed method considers adequate input data information through time series decomposition and adopts an idea of Mixture of Experts(MoE). Time series decomposition allows additional elements such as trend and differencing to be considered from the original time series. MoE consists of expert models and gate models, where the gate controls the weights of each expert model. By employing the idea of MoE, we expect the learning model to better adapt to various input patterns through decomposition and to utilize only necessary experts for reconstruction. Experiments involve comparing the proposed method with performance against more than ten baselines using five benchmark datasets. Based on experimental results, we will discuss the practical potentials.

3 - Noise-Resistant Anomaly Detection in Multivariate Time Series Using Feature Correlation

Hyungkwon Lee, Sungkyunkwan University, Suwon, Korea, Republic of, Jong-Seok Lee

Detecting anomalies within time series data is particularly challenging, especially in domains where anomalies have significant consequences. This task is further complicated by multivariate time series data, which often includes both noisy and noise-free observations due to inherent sensor inaccuracies or operational constraints. The presence of noisy data increases the risk of erroneous anomaly detection, especially in reconstruction-based approaches relying on generative models. To address this issue, we propose a new anomaly detection method based on TadGAN (Time Series Anomaly Detection Using Generative Adversarial Networks). By leveraging the correlation between noise-free and noisy features, our method aims to enhance anomaly detection accuracy and decrease false detection rates. This approach offers a robust solution for detecting anomalies in noisy environments by thoughtfully integrating insights from noise-free data. Numerical experiments demonstrate the effectiveness of our method in various time series data scenarios that include noisy data.

4 - Wavelet scattering for classification in health-related time-series datasets.

Jyotirmoy Nirupam Das, The Pennsylvania State University, State College, PA, United States, Yanling Li, Soundar Kumara, Sy-Miin Chow

Time series data in health-related studies are often characterized by non-stationaries in frequencies over time. Wavelet scattering is a class of machine learning methods that combines the strengths of wavelet transform and deep neural network to enable the use of deep learning for forecasting and classification purposes using features from wavelet analysis across different time and frequency scales within a single step. This work uses wavelet scattering to analyze and extract important features from two health datasets. In the first study, we used wavelet scattering to classify participants' facial electromyography (EMG) activities and self-reports into distinct categorical phases (during baseline, under positive and negative emotion induction procedures, and during recovery). In the second study, we use wavelet scattering on wearable device measures collected from individuals during the early phases of the SARS-CoV2 pandemic to classify individuals' infection status. In both cases, wavelet scattering provided robust feature generation to improve the performance of the deep learning model in discerning the hidden patterns in these health measures.

5 - A robust time series model with support vector regression

Chun-Yen Tsai, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Puli, Taiwan, Jing-Rung Yu, Hao-Feng Deng, Donald Lien

This study incorporates error tolerance from support vector regression (SVR) into the Opt-ARIMA(fw) model, resulting in the SVR_ARIMA(fw) forecasting model tailored for fluctuating data. The Opt-ARIMA(fw) transforms the autoregressive integrated moving average (ARIMA) model using the Haar wavelet function to enhance performance. In addition, to address overfitting and reduce reliance on the subjective judgment in conventional ARIMA, the Opt-ARIMA(fw) simultaneously determines lagged terms, seasonal effect, and coefficients. The resulting SVR_ARIMA(fw) maintains a mixed-integer linear programming structure, inheriting the streamlined benefits of the Opt-ARIMA(fw), enhancing interpretability and efficacy for high-volatility datasets. From our evaluations comparing support vector regression, ARIMA, random forest, and GB, Opt-ARIMA(fw) excels in handling stable datasets but weak in dealing with fluctuating datasets from our preliminary test. Therefore, we employ support vector regression to enhance the Opt-ARIMA(fw), resulting in a more robust forecasting model, SVR_ARIMA(fw) which remains a linear mixed integer programming model to enhance the robustness and accuracy of the Opt-ARIMA(fw) by incorporating error tolerance in addition to automatic determination the seasonal effect and number of lags in MA and AR terms. We expand the training sample size and use the proposed SVR-ARIMA(fw) to forecast non-stationary datasets, such as high-frequency prices or returns of cryptocurrencies and composite stocks of NASDAQ 100 and S&P 500.

6 - Semi-Supervised Time Series Domain Adaptation Based On Contrastive Learning Using Momentum Encoder

Seonyoung Kim, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

Machine learning models perform well on data from a particular domain, such as the data in training time. However, they are degraded when applied to data from a different domain. Many researchers have suggested alleviating these issues, as known as domain shift or discrepancy, in computer vision. Yet, time series has not been extensively studied in comparison to computer vision. These phenomena can be fatal in medical situations, but there are limitations due to the restricted data accessibility and high annotation cost. Likewise, we can easily face the conditions of just a few labeled target data, albeit given sufficient labeled source data in practice. Semi-supervised domain adaptation aims to reduce the discrepancy between the source and target distribution in the presence of limited labeled target samples. This study proposes a semi-supervised time series domain adaptation method using contrastive learning techniques with a momentum encoder to mitigate the domain shifts between the source and target domains and prevent collapse in the model. Additionally, we demonstrate the effectiveness of our method using multivariate bio-medical time series data.

SE67

Regency - 605

George B. Dantzig Dissertation Award

Award Session

George B Dantzig Dissertation Prize Session

Chair: Ruth Beer, Baruch College, CUNY, New York, NY, United States

1 - Simulating Dynamical Systems from Data

Abdullah Alomar, MIT, Oakland, CA, United States

The availability of data from dynamical systems presents opportunities for automated decision-making, but challenges like high dimensionality, noise, sparsity, and confounding persist. We propose methods to leverage these datasets and address these issues in various inference tasks. Central to our approach is a key factorization of the dynamics, shared across related yet heterogeneous systems, into a linear combination of latent separable functions of state and action. This principled structure guides effective methods for time series forecasting, change point detection, reinforcement learning, and trace-driven simulation.

2 - Scalable, Efficient, and Fair Algorithms for Structured Convex Optimization Problems

Mehrdad Ghadiri, Georgia Institute of Technology, Atlanta, GA, United States

The thesis presents algorithms for fundamental optimization problems with theoretical guarantees on approximation quality and running time. These problems include linear programming (LP), tensor decompositions, and fair clustering. The focus is on developing efficient and scalable algorithms with provable approximation guarantees while incorporating ethical considerations such as fairness. Many industry applications motivate these problems. In particular, tensors are building blocks of deep neural networks, the workhorse of large language models (LLMs).

3 - Essays on Technology-enabled Social Impact

Alp Sungu, Wharton, Philadelphia, PA, United States

This thesis explores poverty alleviation policies in developing countries through an operational and technological lens. It sets up a research infrastructure in a low-income settlement in India to run field experiments. The first chapter places research assistants at food vendors to track nutrition for over 20,000 individuals and opens a subsidy program. It finds excessive junk food purchasing, especially among working parents. Staple food subsidies induce a shift from snacking out to home-cooked meals. The second chapter reveals low-income smartphone users binge on social media early in the month and are cut off from digital information later. A daily-capped data plan smooths data consumption and improves access to information.

4 - Analytics and Decision Making in Sustainable Operations

Leann Thayaparan, Johns Hopkins University, Baltimore, MD, United States

As we move towards more renewable resources, the ability to produce electricity in time with demand diminishes. Instead, rises a need for energy storage. Electric Vehicles (EVs) have been discussed as a way of providing a distributed energy storage resource to the electric grid. However, highly complex, non-linear driver behavior must be accounted for. In this dissertation we closely collaborate with a large American EV manufacturer to combine machine learning with optimization to model driver behavior in order to size the capacity of energy storage EVs can offer to the grid.

SE68

Regency - 606

Nicholson Student Paper Competition II

Award Session

Nicholson Student Paper Competition

Chair: Siqian Shen, U of Michigan/NSF, Ann Arbor/Alexandria, 48109, United States

Co-Chair: Andrew Daw, USC Marshall Data Sciences and Operations, Los Angeles, CA, United States

1 - The Role of Level-Set Geometry on the Performance of PDHG for Conic Linear Optimization

Zikai Xiong, MIT OR Center, Cambridge, MA, United States

We analyze the performance of the restarted primal-dual hybrid gradient method (rPDHG) for general conic linear optimization problems, at the scale where matrix-factorization-free methods are attractive or necessary. We define three geometric condition numbers of the primal-dual level set: the diameter, the conic radius, and the Hausdorff distance to the optimal solutions, and we show how these condition numbers inform the convergence rate of rPDHG both in theory and in practice. We further show how central-path-based conic rescaling can markedly enhance the convergence rate of rPDHG.

2 - Optimizing Inventory Placement for a Downstream Online Matching Problem

Boris Epstein, Columbia University, New York, NY, United States

We study the inventory placement problem of splitting Q units of a single item across warehouses, in advance of an online matching problem. We compare the performance of two placement procedures based on optimizing surrogate functions: Offline and Fluid placement. We show that optimizing inventory placement for the Offline surrogate leads to an $\alpha (1-(1-1/d)^d)$-approximation for the joint placement and fulfillment problem under any demand model that admits an α-competitive fulfillment policy. We assume d is an upper bound on how many warehouses can serve any demand location.

3 - Causal Message Passing: A Method for Experiments with Unknown and General Network Interference

Mohamad Sadegh Shirani Faradonbeh, Stanford University, Stanford, CA, United States

Randomized experiments are a powerful methodology for data-driven evaluation of decisions. Yet, their validity may be undermined by network interference. This study introduces a new framework to accommodate complex and unknown network interference. Our framework, termed causal message-passing, is grounded in approximate message passing methodology. Utilizing causal message-passing, we introduce a practical algorithm to estimate the total treatment effect. We demonstrate the effectiveness of this approach across five numerical scenarios, each characterized by a distinct interference structure.

4 - The Cost of Impatience in Dynamic Matching: Scaling Laws and Operating Regimes

Angela Kohlenberg, Northwestern University, Evanston, IL, United States

We study two-sided matching queues with abandonment. We identify universal, non-asymptotic scaling laws for the matching loss due to abandonment (the “cost-of-impatience”). The scaling laws characterize how this cost depends on the arrival rates and mean patience of each side, and reveal four operating regimes. The scaling laws, specialized to each regime, reveal the fundamental structure of the cost-of-impatience and show that it is fully determined by (i) a “winner-take-all” competition between mean patience and utilization, and (ii) the ability to accumulate inventory on one side.

SE69

Regency - 607

The Economics of eBusiness

Invited Session

eBusiness

Chair: Jisu Cao, Arizona State University, Tempe, AZ, United States

1 - The Effect of Law Enforcement's Participation on Online Platforms on User Engagement in Sharing Contributions

Jingyan Dai, University of Florida, Gainesville, FL, United States, Qili Wang, Shu He, Liangfei Qiu

Social media is being used by government organizations more and more to engage with the public they serve. It has rapidly become a significant focal point of government initiatives, quickly becoming an integral part of e-government efforts. The interactions between government and individuals could result in the broadening of government services, the emergence of new ideas, and improvements in decision-making and problem-solving capabilities (Bertot 2012). Nevertheless, increased involvement of government agencies in online activities could potentially alter the behavior of existing online users, particularly within communities where voluntary knowledge contribution is prevalent. To capture the impact of government involvement on users' voluntary knowledge contribution, we narrow our focus to the impact of the access of law enforcement to an online platform. The goal of connecting residents with public safety agencies through online communities is to cultivate safer and more well-informed communities. In this study, we investigate how law enforcement participation on online platforms affects users' voluntary contributions. Our results indicate that when a police agency joins the platform, users tend to decrease their voluntary posting activity. This cautious behavior likely leads to an improvement in the quality of online posts.

2 - Exploring Ar Interactions in the Industrial Metaverse: Effects on Information Search Efficiency and Learning Effectiveness

Runge Zhu, Tsinghua University, Beijing, China, People's Republic of, Ting Li, Cheng Yi

This study ventures into the industrial Metaverse to unravel the intricacies of Augmented Reality (AR) interactions and their effect on users’ information search efficiency and learning effectiveness. Through a large-scale online experiment and a field experiment involving real-world maintenance mechanics from one of the world’s largest airlines, we examined the differing impacts of multimodal and menu-based AR interactions, while also considering the moderating effects of task type, output multiplicity, and user experience. Our findings reveal that multimodal AR interaction is more efficient for information search, particularly for tasks with a higher search depth and less experienced users. However, this effect declines when multiple outputs are present. Regarding learning effectiveness, multimodal interaction hinders learning for tasks with a higher search depth and for experienced users, but facilitates learning when multiple outputs are available. These findings provide crucial insights for AR interaction designs in the emerging Metaverse, with significant implications for industrial training programs.

3 - Bias in Generative AI

Mi Zhou, University of British Columbia, Vancouver, BC, Canada, Vibhanshu Abhishek, Timothy Derdenger, Jaymo Kim, Kannan Srinivasan

This study analyzed images generated by three popular generative artificial intelligence (AI) tools - Midjourney, Stable Diffusion, and DALLE 2 - representing various occupations to investigate potential bias in AI generators. Our analysis revealed two overarching areas of concern in these AI generators, including (1) systematic gender and racial biases, and (2) subtle biases in facial expressions and appearances. Firstly, we found that all three AI generators exhibited bias against women and African Americans. Moreover, we found that the evident gender and racial biases uncovered in our analysis were even more pronounced than the status quo when compared to labor force statistics or Google images, intensifying the harmful biases we are actively striving to rectify in our society. Secondly, our study uncovered more nuanced prejudices in the portrayal of emotions and appearances. For example, women were depicted as younger with more smiles and happiness, while men were depicted as older with more neutral expressions and anger, posing a risk that generative AI models may unintentionally depict women as more submissive and less competent than men. Such nuanced biases, by their less overt nature, might be more problematic as they can permeate perceptions unconsciously and may be more difficult to rectify. Although the extent of bias varied depending on the model, the direction of bias remained consistent in both commercial and open-source AI generators. As these tools become commonplace, our study highlights the urgency to identify and mitigate various biases in generative AI, reinforcing the commitment for a more inclusive future.

4 - Learning to Adopt Generative AI

Lijia Ma, University of Washington, Seattle, WA, United States, Xingchen Xu, Yumei He, Yong Tan

ChatGPT, owing to its versatility, has garnered a substantial user base and has become a staple in many individuals' daily routines. Nonetheless, upon its introduction, potential adopters often face uncertainty regarding the technology. By accessing public information and engaging directly with ChatGPT, users can progressively discern its distinctions from traditional tools such as Google Search, thereby facilitating more informed decisions. However, individual differences—such as educational backgrounds and professional fields—may affect the speed at which users adapt, potentially impeding the adoption process and exacerbating the digital divide. Furthermore, users may derive varied utilities from the same tool based on their unique needs and contexts. To disentangle these mechanisms, we develope a comprehensive Bayesian Learning model incorporating multi-layer heterogeneities and calibrate it using detailed clickstream data. Utilizing the results from this model, we also crafte counterfactual policies aimed at enhancing user welfare.

5 - Bytes and Bets: How Online spORts Gamblers Navigate Risk and Return Through Online Forums

Jisu Cao, Arizona State University, Tempe, AZ, United States

Online forums, a virtual space where individuals can connect and share knowledge, have become a popular information source for users. While these forums provide valuable insights to help users make informed decisions, the quality of the content can vary significantly. Trusting misleading information can negatively impact forum participants, potentially leading to financial losses. In this paper, we try to understand how bettors seek information from online forums and its impact on their risk attitudes and returns. Collaborating with a leading online sports betting company in the United States, we collected a random sample of gamblers' web browsing history and their betting behaviors from 2019 to 2022 and conducted a set of empirical analyses. Our results show an increased risk attitude after the bettor participates in online forums. Specifically, bettors increase the amount of money they wager by 8\% for each additional post they read from the online forum. Surprisingly, these information-seeking behaviors do not improve betting outcomes but hurt bettors' financial returns. For every additional post read by the bettor, the net return decreases by $4.81. Our research underscores the importance of governance in online forums and discusses relevant managerial insights.

SE70

Regency - 701

Operations Research Approaches to Wildfire Prevention and Management

Invited Session

Energy: Natural Resources

Chair: Tomas Lagos, The University of Sydney Business School, Sydney, 2000, Australia

1 - Fuel Treatment Planning Under Uncertainty Using a Bilevel Optimization Framework

Tomas Lagos, The University of Sydney, Sydney, Australia, Nam Ho-Nguyen, Dmytro Matsypura, Oleg Prokopyev

In this work, we develop a framework to minimize bushfires burned area resulting from adversarial attacks through strategic fuel treatment planning. Predicting fire ignition distribution is inaccurate due to the complexity of weather phenomena and the limited informativeness of historical data. Consequently, conservative approaches are sensible for protecting the landscape against worst-case scenarios, thereby ensuring baseline performance. We introduce a bilevel mixed-integer formulation with integer variables at both the upper and lower levels. For the general case, we propose an exact algorithm based on constraint generation, while under specific conditions, we propose a polynomial size MIP formulation that can solve the problem exactly. Through a real case study, we demonstrate the scalability and performance of our approaches, comparing them against heuristic procedures from existing literature.

2 - A Stochastic Programming Rolling Horizon Approach for Extended Attack for Extreme Wildfires

Lina Villa, Texas A&M University, College Station, TX, United States, Lewis Ntaimo

Wildfires pose significant threats to both human lives and the environment. In this study, we propose a pragmatic approach to wildfire management involving dynamic firefighting resource allocation under uncertainty for extreme wildfires. Our approach integrates a two-stage stochastic mixed-integer programming model with a rolling horizon strategy to dynamically allocate firefighting resources to a large-scale wildfire over time. We consider the construction of firelines, whereby combustible fuels are cleared by specific resources (in our case study, dozers), aimed at creating firebreaks to hinder fire growth. A fire manager identifies locations for fireline construction, and the model allocates resources to different points around the fire for fireline construction. The resources are dynamically reallocated at each stage. Decisions from the previous stage are incorporated into a fire behavior simulator to reflect the impact of the fireline construction at the next stage. Stochasticity is incorporated through weather scenarios, which in addition to resource allocation and fireline construction decisions, influence wildfire behavior. Therefore, we adopt a simulation-optimization approach within the rolling horizon framework. Through computational experiments conducted on simulated scenarios with real data, we demonstrate the efficacy of our approach in significantly reducing expected fire spread and impacts through optimized firefighting resource allocation.

3 - Forest Landscape Compartmentalization to Reduce Wildfire Risk

Denys Yemshanov, Natural Resources Canada, Sault Ste. Marie, ON, Canada, Ning Liu, Eric Neilson, Frank Koch

Wildfires is a dominant disturbance agent in North-American forests. A strategic removal of forest fuels to create landscape compartments has been deemed a viable strategy to reduce the wildfire spread risk. However, limited resources make it challenging to plan compartmentalization effectively. We consider a landscape compartmentalization problem to minimize the risk of wildfires under a resource constraint. We depict a forest landscape as a network of patches with flammable fuels and formulate a Critical Edge Removal problem to find a configuration of firebreaks that minimizes the risk of wildfire spread in the area of concern. We use a spatial fire growth model to generate a set of fire spread likelihoods between all pairs of locations in the landscape and use this set to guide the allocation of firebreaks. We compare the proposed problem with the critical node detection problem that removes a subset of nodes from a connected landscape fuel network to minimize its connectivity. We apply the approach to assist with the wildfire risk mitigation planning in the Red Rock-Prairie Creek area, Alberta, a complex, fire-prone landscape. Our solutions provide several strategies for reducing wildfire hazard in fire-prone landscapes and could assist strategic planning of wildfire mitigation activities in other regions.

SE71

Regency - 702

Machine Learning for Optimization

Invited Session

Data Mining

Chair: Jay Rosenberger, University of Texas-Arlington, P.O. Box 19017, Arlington, TX, 76019, United States

Co-Chair: Yuan Zhou, University of Texas at Arlington, Arlington, TX, United States

1 - Developing a Composite Objective Function for Evaluating Chromatography Extraction Quality

Jaivardhan Sood, University of Texas at Arlington, Arlington, TX, United States, Jay Rosenberger, Victoria Chen, Kevin Schug, Niray Bhakta, Srividya Sekar

This research focuses on determining the weights associated with parameters that comprehensively capture the extraction science of a compound. We aim to create a composite objective function to evaluate compound extraction quality during chromatography, moving beyond single-metric assessments. Utilizing the Analytical Hierarchy Process (AHP) and linear programming, we derive these weights, which also account for the hierarchy of runs as evaluated by chemists. This method provides a holistic assessment tool that enhances decision-making in analytical chemistry. Our approach ensures a well-rounded evaluation of extraction processes, potentially improving the accuracy and reliability of compound extraction analysis in chromatography, thereby supporting better outcomes in analytical chemistry practices. Additionally, this composite function will be used in conjunction with the global optimization methodology QMARS-SUROPT to optimize parameters for SFS chemistry equipment, further enhancing the effectiveness of the extraction process.

2 - An Adaptive Framework for Continuous-State Stochastic Dynamic Programming

Huiyuan Fan, The University of Texas at Arlington, Arlington, TX, United States, Prashant Tarun, Victoria Chen

This presentation introduces a sequential state space exploration (SSSE) approach for approximate dynamic programming (ADP) inspired by reinforcement learning. The SSSE approach is coupled with an adaptive value function approximation (AVFA) algorithm that gradually grows the complexity of the statistical model as more data are observed. We refer to the novel SSSE-AVFA approach as an adaptive dynamic programming algorithm, as it automatically identifies the appropriate state space, sample size, and statistical model structure, thus, eliminating the need for time-consuming trial-and-error computational runs that were previously required. The new SSSE-AVFA algorithm is demonstrated using a nine-dimensional inventory forecasting problem and is compared with fixed structure runs in which the state space, sample size, and statistical model structure are assumed in advance. The implementation presented here employs neural networks and quasi-random, low-discrepancy sequence experimental designs. A single run of SSSE-AVFA algorithm takes longer than a single run of a fixed structure method. However, it must be recognized that if the SSSE-AVFA algorithm takes, for example, 20 times longer, that is still a significant improvement over many, many trial-and-error runs, as such a trial-and-error process could very easily go over 20 runs. As the newly proposed SSSE-AVFA algorithm is adaptive in determining the state space, sample size and statistical model structure, it has the potential to be more effective and efficient than the traditional methods in handling a wide range of real-world DP problems.

3 - Approximate Dynamic Programming with A Relu-Neuralnet for Temporary Healthcare Facilities

Vigneshwar Pesaru, University of Texas at Arlington, Arlington, TX, United States, Jay Rosenberger

The outbreak of coronavirus disease 2019 (COVID-19) has seriously affected the whole world, and pandemic research has attracted increasing amounts of scholarly attention. For instance, inappropriate patient allocation may result in unequal access and/or higher service costs in different regions at different time periods. Therefore, it is crucial to integrate capacity planning in healthcare system design while taking into consideration dynamic population requirements and uncertainties in patient demand, attrition, and death rates. In this research, we construct a mathematical model for a temporary healthcare facility problem that considers uncertainties in demand and attrition along with transferring patients between hospitals. We implement a future value approximation method with a ReLU-NeuralNet. We adopt a stopping criterion to effectively identify an optimally equivalent value function. Finally, we conduct numerical experiments to investigate the performance of the proposed method.

4 - Predicting Gas Phase Absorption Spectra in the Vacuum Ultraviolet Range through Molecular Feature Representations and Machine Learning Methods

Matine Ghasemloo, University of Texas at Arlington, Arlington, TX, United States, Victoria Chen, Jay Rosenberger, Kevin Schug, Shouyi Wang, Yujing Yang, Linh Ho Manh

Ultraviolet (UV) absorption spectroscopy stands as a widely employed technique for both quantitative and qualitative analysis of chemical compounds. Particularly in the gas phase, vacuum UV (VUV) and UV absorption spectra offer specific and diagnostic insights into numerous small molecules. Precisely predicting VUV/UV absorption spectra holds significant potential for characterizing novel or unidentified molecules across diverse fields such as fuels, forensics, and pharmaceutical research. Artificial intelligence emerges as a promising alternative to conventional quantum chemical methods for spectral prediction. In this study, diverse molecular feature representation techniques were explored and refined to encode chemical structures, facilitating the evaluation of three distinct machine learning models for predicting gas phase VUV/UV absorption spectra. Utilizing structure data files (.sdf) and VUV/UV absorption spectra from 1397 volatile and semi-volatile chemical compounds, the models were trained and tested. Novel molecular features (referred to as ABOCH) were introduced to better encapsulate pi-bonding, aromaticity, and halogenation effects. Incorporating these new features significantly enhanced spectral prediction accuracy, outperforming computationally-intensive molecular-based deep learning approaches. Among the machine learning methods examined, the Random Forest regressor demonstrated the highest accuracy score with the shortest training duration. Moreover, the developed machine learning prediction model surpassed spectral predictions relying on time-dependent density functional theory, underscoring its efficacy and practicality.

SE72

Regency - 703

Roundoff-Error-Free Solutions of Optimization Problems

Invited Session

Computing Society

Chair: Christopher Lourenco, US Naval Academy, Annapolis, MD, United States

1 - Expediting Exact LP Solvers

Christopher Lourenco, US Naval Academy, Annapolis, MD, United States

Linear programming is the foundation of mathematical optimization; however, more than 20 years of research has shown that nontrivial round-off-errors can lead to incorrect solutions on a sizeable subset of real-world LPs. Currently, when exact or higher precision solutions are needed, the state-of-the-art LP solver is SoPlex which guarantees exactness via rational-arithmetic LU factorizations. This work accelerates the exact LP solver by replacing these rational factorizations and performing all KKT conditions via the exact LU factorizations of the SPEX software package. Altogether, the new interface increases the speed of SoPlex by a factor of 2.5 and 1.7 in average and geometric mean on a test set of real-world linear programs.

2 - Exact QR factorization

Erick Moreno-Centeno, Texas A&M University, College Station, TX, United States

We present the roundoff-error-free (REF) QR factorization framework comprising integer-preserving versions of the standard and the thin QR factorizations and associated algorithms to compute them. Specifically, the standard REF QR factorization factors a given matrix $A \in Z^{m \times n}$ as A = QDR, where $Q \in Z^{m \times m}$ has pairwise orthogonal columns, D is a diagonal matrix, and $R \in Z^{m \times n}$ is an upper trapezoidal matrix; notably, the entries of Q and R are integral, while the entries of D are reciprocals of integers. In the thin REF QR factorization, $Q \in Z^{m \times n}$ also has pairwise orthogonal columns, and $R \in Z^{n \times n}$ is also an upper triangular matrix. In contrast to traditional (i.e., floating-point) QR factorizations, every operation used to compute these factors is integral; thus, REF QR is guaranteed to be an exact orthogonal decomposition. Importantly, the bit-length of every entry in the REF QR factorizations (and within the algorithms to compute them) is bounded polynomially. Notable applications of our REF QR factorizations include finding exact least squares or exact basic solutions, $x \in Q^n$, to any given full column rank or rank deficient linear system Ax=b, respectively. In addition, our exact factorizations can be used as a subroutine within exact and/or high-precision quadratic programming. Altogether, REF QR provides a framework to obtain exact orthogonal factorizations of any rational matrix (as any rational/decimal matrix can be easily transformed into an integral matrix).

3 - Sparse Exact QR Factorization

Lorena Mejia Domenzain, Texas A&M University, College Station, TX, United States

The Gram-Schmidt process is most widely used for computing the QR factorization of a matrix. QR factorization is a widely used tool in mathematics, computer science, and engineering. For example, QR factorization is regularly used to solve least squares problems. Round-off Error Free (REF) QR was developed because the round-off errors incurred when factorizing can lead to a loss in orthogonality (thus incorrect solutions). REF QR factorization of matrix A is A=QDR, where Q is a left orthogonal matrix, R is a right triangular matrix, and D is a diagonal matrix. This talk presents an Integer-preserving Gram-Schmidt orthogonalization process (IPGS) and an efficient sparse IPGS-REF QR algorithm. IPGS is strongly grounded in exact linear algebra and every operation used in IPGS is integral; thus, the resulting vectors are guaranteed to be exact, integral and orthogonal to each other.

4 - An Early Termination Criteria For Dixon's Method

Kelsey Kitzmiller, Texas A&M University, College Station, TX, United States

Dixon’s p-adic lifting algorithm uses modular arithmetic to find exact solutions to systems of linear equations. Although this method is state-of-the-art for solving dense systems exactly, it often performs significantly more iterations than necessary to compute the solution. In order to avoid some of these unnecessary computations, a new early termination criteria is presented which is based on Cabay’s criteria for congruence technique algorithms. This stopping procedure is predicated on a new version of rational reconstruction which partially reconstructs the modular image found at each p-adic lifting iteration. The complete algorithm maintains the same worst-case complexity as Dixon’s original method but is more efficient in practice, as it performs only a small number of iterations beyond those which are necessary to compute the solution.

SE73

Regency - 704

Statistical and Machine Learning in Healthcare

Invited Session

Data Mining

Chair: Xin Zan, University of Iowa, Coralville, IA, United States

1 - Muse-Net: Missingness-Aware Multi-Branching Self-Attention Encoder for Irregular Longitudinal Electronic Health Records

Zekai Wang, The University of Tennessee. Knoxville, Knoxville, TN, United States, Tieming Liu, Bing Yao

The era of big data has made vast amounts of clinical data readily available, particularly in the form of electronic health records (EHRs), which provides unprecedented opportunities for developing data-driven diagnostic tools to enhance clinical decision making. However, the application of EHRs in data-driven modeling faces challenges such as irregularly spaced multi-variate time series, issues of incompleteness, and data imbalance. Realizing the full data potential of EHRs hinges on the development of advanced analytical models. In this paper, we propose a novel Missingness-aware mUlti-branching Self-Attention Encoder (MUSE-Net) to cope with the challenges in modeling longitudinal EHRs for data-driven disease prediction. The MUSE-Net leverages a multi-task Gaussian process (MGP) with GPU acceleration for data imputation, a multi-branching architecture to address the data imbalance problem, and a time-aware self-attention encoder to account for the irregularly spaced time interval in longitudinal EHRs. We evaluate the proposed MUSE-Net using both synthetic and real-world datasets. Experimental results show that our MUSE-Net outperforms existing methods that are widely used to investigate longitudinal signals.

2 - Detecting Algorithmic Bias in Medical-AI Models Using Conformal Trees

Jeffrey Smith, Georgia Tech, Atlanta, GA, United States

With the growing prevalence of machine learning and artificial intelligence-based medical decision support systems, it is equally important to ensure that these systems provide patient outcomes in a fair and equitable fashion. This study presents an innovative framework for detecting areas of algorithmic bias in medical-AI decision support systems. Our approach efficiently identifies potential biases in medical-AI models, specifically in the context of sepsis prediction, by employing the Classification and Regression Trees (CART) algorithm with conformity scores. We verify our methodology by conducting a series of synthetic data experiments, showcasing its ability to estimate areas of bias in controlled settings precisely. The effectiveness of the concept is further validated by experiments using electronic medical records from Grady Memorial Hospital in Atlanta, Georgia. These tests demonstrate the practical implementation of our strategy in a clinical environment, where it can function as a vital instrument for guaranteeing fairness and equity in AI-based medical decisions.

3 - Heterogeneous Elasticity Moduli Estimation Based on Bone Mechanical Testing Data by Semi-Parametric Physics-Informed Neural Networks

Tatthapong Srikitrungruang, Texas A&M University, College Station, TX, United States, Sina Aghaee Dabaghan Fard, Jaesung Lee, Yuxiao Zhou

Estimation of bone stiffness is challenging, some research done previously uses physical experiments such as microtensile testing which is expensive and destructive. Simulation is also an option. But, it is still expensive and more suitable for response prediction whereas parameters estimation is limited. Physics-informed neural networks (PINNs) merge data-driven methodologies with laws of physics and have some effectiveness in parameter estimation. However, the estimation of heterogeneous physical parameters such as elasticity moduli is very challenging because one needs to estimate the parameter value at every given input location. Moreover, It is more challenging in three-dimensional space. In practice, we often have data of additional variables that have parametric relationships with heterogeneous parameters. For example, the parameter relations between bone stiffness and the bone mineral density measured from computed tomography (CT) data is studied in the literature. We leverage availability of such a variable and the parametric relationships by integrating them into PINNs framework. A novel Semi-parametric Physics-Informed Neural Networks is proposed to address challenges in optimization for parameter estimation. Our framework consists of two steps; in the first step, we optimize the objective function specially designed for the inverse linear elasticity model, where we incorporate slack variables to address noises in the data. Instead of using the boundary condition that exacerbates the optimization, we first find an unscaled parameter and rescale it. We demonstrate heterogeneous elastic moduli estimation based on the glenoid bone displacement data from mechanical testing coupled with CT images, and the result corresponds to the literature.

4 - A Dual-Granularity Bayesian Active Learning Framework with Uncertainty Quantification for Sleep Apnea Severity Estimation

Xin Zan, University of Florida, Gainesville, FL, United States

This work develops a dual-granularity Bayesian active learning framework with uncertainty quantification for automatically providing more reliable and accurate fine-grained apnea severity estimation. The proposed novel framework provides efficient manual annotations without remarkable diagnosis accuracy degradation by using available coarse-grained labeled data and actively querying instant-level labels at the fine granularity. In particular, the proposed framework employs a Bayesian deep learning architecture to systematically quantify the irreducible aleatoric uncertainty and reducible epistemic uncertainty embedded in acquired signal data and unknown model parameters, respectively. A novel integrated acquisition function, which combines the quantified epistemic uncertainty with knowledge misalignment, is designed to facilitate selecting the instants whose estimation contradicts clinical knowledge and with high epistemic uncertainty in instant-level active learning. A dual-granularity negative Gaussian log likelihood loss function is also designed and seamlessly integrated into the active learning framework to unify the model training process from iteratively updated training dataset with both initial coarse-grained labels and the additional queried fine-grained labels. An experiment is conducted using a real sleep dataset to demonstrate the superior accurate and reliable fine-grained apnea severity estimation of the proposed method.

5 - A Mutual Knowledge Distillation Framework for Alzheimer&Rsquo;S Disease Diagnosis Using Incomplete Multimodal Images

Min Gu Kwak, Georgia Institute of Technology, Atlanta, GA, United States, Lingchao Mao, Zhiyang Zheng, Yi Su, Fleming Lure, Jing Li

Early detection of Alzheimer's Disease (AD) is crucial for timely interventions and optimizing treatment outcomes. Despite the promise of integrating multimodal neuroimages such as MRI and PET, handling datasets with incomplete modalities remains under-researched. This phenomenon, however, is common in real-world scenarios as not every patient has all modalities due to practical constraints such as cost, access, and safety concerns. We propose a deep learning framework employing Mutual Knowledge Distillation (MKD) to model different sub-cohorts of patients based on their available modalities. In MKD, the multi-modality model (e.g., MRI and PET) serves as a teacher, while the single-modality model (e.g., MRI only) is the student. Our MKD framework features three components: a Student-oriented Multimodal Teacher (SMT) model designed through information disentanglement, a student model that learns from classification errors and SMT's knowledge, and the teacher model enhanced via distilling the student's single-modal feature extraction capabilities. Moreover, we show the effectiveness of the proposed method through theoretical analysis and validate its performance with simulation studies. In addition, our method is demonstrated through a case study with Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets, underscoring the potential of artificial intelligence in addressing incomplete multimodal neuroimaging datasets and advancing early AD detection.

SE74

Regency - 705

Optimization and Data Analytics for Agriculture Operations and Biomass Supply Chains

Invited Session

ENRE: Environment and Sustainability

Chair: Neng Fan, University of Arizona, Tucson

1 - Viability of Distributed Manufacturing for Chemicals from Biomass Using Local Renewable-Powered Electrochemistry

Sarah Ryan, Iowa State University, Ames, IA, United States, Motahareh Kashanian, Shayan Tohidi

A hybrid microbial electrochemical system (HMES) is a form of process intensification for converting biomass into chemicals, in which electrochemical reactions are done directly in the fermentation broth. Electrochemical reactors lack the economies of scale that characterize many types of chemical process equipment. This consideration, when combined with the cost of transporting feedstock and the potential for using local renewable energy, suggests that small-scale, spatially dispersed production could be economically and environmentally sustainable. We investigate this possibility in a two-step procedure to select sizes and locations of production facilities that minimize cost, including the social cost of carbon emissions. First, we use the REopt tool from the National Renewable Energy Laboratory to optimize capacities of wind and solar generation, at each potential facility location and size. Second, we solve a mixed-integer program to identify the optimal facility locations and sizes to meet a given demand for chemicals. Through detailed sensitivity analyses, we identify the economic, technical, and environmental factors that have the most significant impacts on the viability of distributed manufacturing for HMES. Lower economies of scale for the non-electrochemical processes, longer plant life and higher estimates of the social cost of carbon encourage a more distributed configuration of production facilities.

2 - Crop Yield Prediction for Semi-Arid Regions

Mahdi Mahdavimanshadi, University of Arizona, TUCSON, AZ, United States, Neng Fan

Predicting crop yields is complex due to various factors like crop genotype, environmental conditions, management practice, and their correlations. This research uses a deep learning approach combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to forecast crop yields using environmental, soil, and management data. The proposed CNN-RNN model was compared with traditional methods like random forest (RF), deep fully connected neural networks (DFNN), SVM, and LASSO using historical data of wheat and cotton from Arizona for 1990-2023. Results show that the CNN-RNN model achieved significantly lower RMSE than other methods, indicating its superiority in yield prediction. In this paper, we demonstrate consistency in yield prediction across diverse environments and provide insights into the influence of weather, soil, and management practices on yield variability, especially in semi-arid regions compared to other areas.

3 - Identifying Sub-Communities in Intestinal Microbiome of Chickens

Mohammad Fili, Oklahoma State University, Stillwater, OK, United States, Guiping Hu, Lizhi Wang, Glenn Zhang

Eimeria maxima is a primary cause of coccidiosis in chickens and a significant predisposing factor for other economically important diseases, such as necrotic enteritis. Despite its impact, the intestinal microbiome's response to E. maxima infection remains inadequately understood. This study aimed to identify the sub-communities within the intestinal microbiome over a 14-day period following E. maxima infection. Bacterial 16S rRNA gene sequencing was conducted on ileal and cecal digesta samples collected from both mock-infected and E. maxima-infected chickens at various infection phases: prepatent (3 days post-infection [dpi]), acute (5 and 7 dpi), and recovery (10 and 14 dpi). By analyzing the abundance trajectories of the most frequent amplicon sequence variants (ASVs), we identified distinct sub-communities within the intestinal microbiome.

4 - Nationwide Optimization of Oilseed Supply Chain for Sustainable Aviation Fuel Production Under Uncertainties

Juliana Pin, North Carolina State University, Raleigh, NC, United States, Daniela Jones, Damon Hartley, Pralhad Burli, Matthew Langholtz, Chad Hellwinckel, David Thompson

The transition to sustainable aviation fuel (SAF) derived from non-petroleum sources is critical for reducing the environmental impact of aviation. The U.S. aims for 35 billion gallons per year (BGY) by 2050. Current studies focus on regional oilseed supply systems. Our research expands this by analyzing the nationwide feasibility of winter oilseeds—pennycress, camelina, and carinata—as viable feedstocks, identifying strategic locations for processing facilities and biorefineries to minimize costs. Using crop production predictions and a Mixed Integer Linear Programming model, we assess potential production volumes, costs, and logistical challenges. If all oilseed-producing counties contribute to SAF production by 2048, 1.12 billion gallons of SAF could be generated, achieving only 4% of the 2050 target at a biorefinery gate price of $0.68 per liter of bio-oil. This highlights the need to diversify feedstock sources. Also, considering meal credit as a co-product is crucial, otherwise costs could increase by 64-70%. We also modeled specific SAF demand levels for 2048 at 350, 700, and 1,050 million gallons—1%, 2%, and 3% of the 2050 target, respectively, with estimated bio-oil unit costs of $0.50, $0.57, and $0.65 per liter at the biorefinery gate. Our analysis addresses both the technical and economic aspects of SAF production and emphasizes the importance of scaling up infrastructure and exploring additional feedstocks to efficiently achieve the U.S. SAF production targets. Building on these results, our current work will enhance model flexibility with multi-capacity processors and biorefineries and include a stochastic model to incorporate uncertainties in oilseed production and SAF demand.

5 - Optimal Rice Cropping Patterns: Synergies in Water and Land Resources Management and Greenhouse Gas Reduction

Shiqi Tan, Sichuan University, Business School, Chengdu, China, People's Republic of, Jian Li, Liming Yao

This study examines optimal rice cropping patterns that effectively manage water and land resources and reduce CH_4 emissions. Rice cultivation stands as a primary source of methane emissions. Water-saving irrigation and optimized resource allocation are crucial for enhancing rice yield and quality while simultaneously reducing greenhouse gas emissions. However, achieving an economically viable yield from rice production is complex, influenced by market demands, constraints on land resources, and policy limitations. These factors form a dynamic decision-making environment. We study a two-stage stochastic dynamic optimization model that incorporates water rights and land allocation, and accounts for the uncertainties of rainfall and market prices, introducing economic incentives for methane reduction.

SE75

Regency - 706

Industry-Research Analysis Partnerships to Inform Energy Infrastructure Decisions

Invited Session

ENRE: Energy-Climate

Chair: Stuart Cohen, National Renewable Energy Laboratory, ARVADA, CO, United States

1 - Hydropower Flexibility and Environmental Tradeoffs Analysis

Thushara De Silva, National Renewable Energy Laboratory, Golden, CO, United States

The importance of hydropower increases as the power grid evolves with the higher variable renewable contribution. As conventional thermal power plants are retired, the importance of hydropower contribution increases to balance the variability of solar and wind generation. However, reservoir water resources are constrained by multiple constraints, and variability of water inflow to the reservoirs creates limitations to dam water releases for power grid needs. Coordinating multiple tools, including water resources, ecological, and technical and economic power grid modeling, informs dam water releases. The case study, the Columbia River Basin multipurpose reservoir project, is operated for hydropower production and many other purposes considering the aquatic habitat of the river basin. Specifically, the river basin fish population is a vital element for the tribal community of the river basin. We integrated a production cost model, a water resource model, and decades of tribal knowledge to analyze the fish-friendly way of operating Columbia hydropower scheduling and grid impacts. We measure power grid impacts for various water resources planning scenarios in terms of total system operating cost, system reliability indicators, changes in wind and solar generation and curtailments, local marginal prices, and revenue for hydropower producers. The study results inform reservoir operating rules decisions from hydropower power producers, system operators, other water users, tribes, environmentalists, and other stakeholders.

2 - A Utility-National Lab Partnership to Value Pumped Storage Hydropower in an Evolving Grid

Stuart Cohen, National Renewable Energy Laboratory, Arvada, CO, United States

Accelerating deployment of variable renewable electricity from wind and solar technologies has created renewed interest in pumped storage hydropower (PSH) as an energy storage technology that can balance energy and maintain grid reliability and resiliency. The Tennessee Valley Authority (TVA) is an electric utility particularly interested in expanding its PSH fleet, but project advancement is challenged by high capital costs relative to internal economic value assessments using available market and analysis tools. That is why TVA partnered with three U.S. Department of Energy National Laboratories (NREL, ANL, and PNNL) to expand upon its internal PSH valuation studies by utilizing state-of-the-art lab capabilities and expertise in hydrological and electricity modeling. This presentation focuses on how the National Renewable Energy Laboratory (NREL) is using its suite of electric grid planning, operations, resource adequacy, and stability modeling tools to better understand the range of potential PSH system values under an uncertain and evolving grid. It will describe how laboratory capabilities are being adapted to a practical utility context and demonstrate preliminary results of PSH value and how it compares with TVA assessments.

3 - Extracting Insight from Unstructured Solar Operations and Maintenance Text Logs

Charity Sotero, kWh Analytics, Beaverton, OR, United States

As solar energy adoption grows, minimizing costly failures is critical. This presentation will detail a Department of Energy-funded analysis of unsupervised text data of over 60,000 O&M tickets from more than 700 U.S. utility-scale solar systems to identify trends in failure-driving equipment and possible resolutions. Natural language processing was used to categorize these unsupervised tickets, with an accuracy rate of 86.2% (F1=0.87). Analysis of the text data found that the top failures occurred in inverters (64.8%), modules (20.9%), trackers (4.7%), and transformers (2.0%). Root causes were often determined to originate much earlier in the PV system for failures originally associated with AC-side equipment. Our findings demand action with implications for each sector of the PV industry: increased on-site parts inventory, proactive maintenance protocols, and data transparency between operators and OEMs to pinpoint and prevent failures before they strike. This knowledge has further implications for PV lenders, insurers, and asset owners to move towards a "safe driver discount" for responsible developers. Overall, this research demonstrates both the richness of PV text data and the challenges associated with pre-processing O&M logs. Insights derived from cross-fleet analytics can help the industry as whole minimize downtime and improve availability.

SE76

Regency - 707

Energy Systems Modeling and Optimization IV: Emerging and Distributed Energy Systems

Invited Session

ENRE: Other Energy

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, 80401, United States

1 - Analysis of Home Battery Energy Storage Systems for Affordable Housing

Andrew Aikman, Colorado School of Mines, Golden, CO, United States, Alexandra Newman, Paulo Cesar Tabares Velasco, Qiuhua Huang

We seek to evaluate the benefits of stationary and non-stationary home battery energy storage systems (BESS) combined with grid-enhancing solutions at the home level to reduce CO2–equivalent and other marginal emissions and to promote energy equity. BESS can be implemented in diverse and energy-burdened neighborhoods in order to support President Biden's Justice40 goals, in addition to reducing environmental impacts and increasing the flexibility and resilience of electric power systems. These disadvantaged communities are marginalized by underinvestment and overburdened by pollution, leading to worse living conditions and lower quality of life. However, BESS are still cost-prohibitive for most low- and middle-income (LMI) households. Thus, wide-scale BESS adoption requires the right mix of incentives and quantification of benefits for community acceptance.
We utilize NREL's REopt model to determine and compare the economic, environmental, and resiliency metrics of home systems consisting of different numbers of batteries . Increasing the system size allows homes to take advantage of time-of-use pricing schemes and periods of low marginal emissions for a reasonable cost tradeoff. Furthermore, we analyze the power outage durations, frequencies, and minimum critical loads that LMI households and communities are capable of handling. Preliminary results demonstrate that the inclusion of between 1 and 4 batteries per household results in 7% to 19% reduction in CO2 emissions, mitigating up to 35 tons of CO2 per year per community. In addition to life cycle cost comparisons, we explore economic incentives necessary to reach various emissions reductions targets.

2 - X Phase Change Material Heat Exchanger to Decarbonize Water Heating in Low-Income Housing

Chase Robinson, Colorado School of Mines, Golden, CO, United States

This project details the fabrication of a phase change material (PCM) heat exchanger for use in small-volume heat pump water heaters (HPWHs) to increase thermal storage capacity and overall performance. The design of the novel heat exchanger relies on additive manufacturing to produce complex geometries not possible via traditional manufacturing methods. This presentation describes efforts toward material processing and performance characterization for the 3D-printed PCM heat exchanger, as well as preliminary results showing the applicability of additive manufacturing for introducing thermal energy storage into a HPWH. Our PCM selection relies on those with low toxicity, high energy storage density, and an appropriate transition temperature for water heating applications. Based on preliminary HPWH modeling results, our work utilizes a material with a transition temperature of 55°C. A higher material content corresponds to a heat exchanger with a higher thermal energy storage density, thus providing greater benefit for small-volume HPWHs. This work presents a promising a path forward towards achieving superior HPWH performance in space-constrained, low-income housing.

3 - Estimating Resource Adequacy Contribution when Energy Storage Provides Frequency Regulation

HyeongJun Kim, The Ohio State University, Columbus, OH, United States, Ramteen Sioshansi

When energy storage capacity value is estimated, state-of-energy (SOE) is important as it determines availability of the storage to alleviate a loss of load. However, in case of providing frequency regulation, estimating the level of SOE becomes complicated and it may become intractable. Thus, we propose a three-step approach to estimate the capacity value of energy storage when it provides frequency regulation. First, discretized stochastic dynamic optimization is used to obtain optimal decision policy when multiple services are provided together. Second, these decision policies are used iteratively to derive actual optimal schedule for the energy storage operation using mixed integer optimization in a rolling-horizon fashion. As the last step, contribution from energy storage to resource adequacy of the power system is assessed using Monte Carlo simulation. Results of the case study are compared with the case when frequency regulation is not provided from energy storage. It shows that the capacity value decreases when there is no non-performance penalty for the capacity or storage duration is short.

4 - Leveraging Advanced Metering Infrastructure to Evaluate Energy Insecurity Demographics

Akua McLeod, Carnegie Mellon University, Pittsburgh, PA, United States, Destenie Nock, Amritanshu Pandey

Interrupted electricity access negatively impacts household resilience and can cause energy insecurity, by hindering households' ability to meet their energy demand. Households can experience an electric service interruption in the form of a power outage, or as an involuntary disconnection due to nonpayment. Often, electricity reliability is reported through system-level metrics and calculated only for only one type of service loss: power outages. These metrics overlook disconnections as a form of electric service loss, and obscure how energy access varies across the service territory. Advanced metering infrastructure (AMI) records sub-hourly electricity consumption at the household level, providing an opportunity to evaluate both forms of service loss and measure differences in access with greater spatial granularity. In this paper, we leverage extensive data from over 600,000 households in Illinois to build a classification model for disaggregating power outages from disconnections in residential electricity consumption smart meter data. We further evaluate how the duration and frequency of disconnections change across the service territory. Our analysis provides more granular metrics for electricity reliability, looking first at involuntary service loss, as well as the combined burdens of outages and disconnections. Ultimately, this analysis informs the identification of energy-insecure communities.

5 - The influence of regional geophysical resource variability on the value of single- and multi-storage technology portfolio

Anna Li, Princeton University, Princeton, NJ, United States, Edgar Virguez, Jacqueline Dowling, Alicia Wongel, Linda Covelli, Tyler Ruggles, Natasha Reich, Nathan Lewis, Ken Caldeira

A stylized macro-scale energy model of least-cost electricity systems relying only on wind and solar generation was used to assess the value individually and in combination of different storage technologies for the contiguous U.S. as well as for four geographically diverse U.S. load-balancing regions. For the contiguous U.S. system, at current costs, when only one storage technology was deployed, hydrogen energy storage produced the lowest system costs, due to its energy-capacity costs being the lowest of all storage technologies modeled. Additional hypothetical storage technologies were more cost-competitive than hydrogen (long-duration) only at very low energy-capacity costs, but they were more cost-competitive than Li-ion batteries (short-duration storage) at relatively high energy- and power-capacity costs. Moreover, in all load-balancing regions investigated, the least-cost systems that included long-duration storage had sufficient energy and power capacity to also meet short-duration energy and power storage needs, so addition of short-duration storage as a second storage technology did not markedly reduce total system costs. Thus, in electricity systems relying on wind and solar generation, contingent on social and geographic constraints, long-duration storage may be able to cost-effectively provide the services that would otherwise be provided by shorter-duration storage technologies.

SE77

Regency - 708

Emerging Applications and Algorithms in Logistics and Urban Planning

Invited Session

Computing Society

Chair: Yiling Zhang, University of Minnesota, Minneapolis, MN, United States

1 - Accelerating Dynamic Programming via Dual Selection for Column-Generation-Based Exact Solution Framework

Zhengzhong You, University of Florida, Gainesville, FL, United States, Yu Yang

Column generation (CG) is a widely used exact solution approach for solving large-scale optimization problems such as vehicle routing problems (VRPs). However, the pricing subproblem, often formulated as a resource-constrained shortest path problem (RCSPP) and solved using dynamic programming (DP), can be computationally expensive. This paper proposes a novel dual selection strategy to accelerate the single iteration of DP in the context of CG. The authors introduce an abstract optimization program to find the "best" optimal dual solution that potentially minimizes the number of undominated labels in the labeling algorithm, a common DP technique for solving RCSPP. The paper presents a set of dual selection programs based on the abstract optimization program, along with a comprehensive theoretical analysis to determine the proper objective function for each program. Numerical experiments on the Capacitated Vehicle Routing Problem (CVRP) demonstrate that the proposed dual selection programs can eliminate up to a significant portion of the undominated labels, resulting in a substantial speedup of the labeling process, particularly for instances where pricing proves to be extremely challenging.

2 - Mobile Parcel Locker Scheduling with Customer Choices Under Uncertain Demand

Yin Can, University of Minnesota, Twin Cities, Minneapolis, MN, United States, Yiling Zhang

The movable unit equipped with a set of parcel lockers has been recently developed as a new mode to improve the efficiency of last-mile delivery. Mobile parcel lockers, different from traditional facility location problems placing permanent facilities in a specific region, can be relocated at any time, driven by the changing demand rate from location to location over time. This talk will present a scheduling problem of mobile parcel lockers under uncertain demand and customer choices. The problem is formulated as a distributionally robust two-stage stochastic programming model, where customer choices are captured by multinomial logit (MNL) models. The probability distribution of uncertain demand, although unknown for this new service, is characterized by the first moments and the support. Our models flexibly adapt different prior beliefs of spatial structures and pricing strategies. We obtain exact mixed-integer linear programming reformulations and derive valid inequalities used in decomposition algorithms.

3 - An Exact Approach for Solving Pickup-and-Delivery Traveling Salesman Problems with Neighborhoods

Cai Gao, University at Buffalo, San Jose, CA, United States, Ningji Wei, Jose Walteros

This paper studies a variant of the traveling salesman problem, called the pickup-and-delivery traveling salesman problem with neighborhoods, that combines traditional pickup and delivery requirements with the flexibility of visiting the customers at locations within compact neighborhoods of arbitrary shape. We derive two optimality conditions for the problem, a local condition that verifies whether a given tour is locally optimal at the visiting points and a global condition that can be used to cut off suboptimal regions of the neighborhoods. We model the problem as a mixed-integer nonlinear program and propose a generalized Benders decomposition to solve instances of the problem with convex and nonconvex neighborhoods. Finally, we conduct extensive computational experiments to demonstrate the efficacy of our solution framework.

4 - Solving the large-scale On-Demand Multimodal Transit Systems Design with Adoptions (ODMTS-DA) Problem

Hongzhao Guan, Georgia Institute of Technology, Atlanta, GA, United States, Beste Basciftci, Pascal Van Hentenryck

Capturing latent demand has a pivotal role in designing public transit services: omitting these riders can lead to poor quality of service and/or additional costs. This study introduces the On-Demand Multimodal Transit Systems (ODMTS) Design with Adoptions problem (ODMTS-DA) to capture the latent demand in on-demand multimodal transit systems. ODMTS-DA is a bilevel optimization problem that integrates a mixed-integer programming (MIP) model and a travel mode choice model. An exact combinatorial algorithm based on Benders decomposition can be applied to solve the problem. Unfortunately, this algorithm only finds high-quality solutions for medium-sized cities and is not practical for large metropolitan areas. To address these computational difficulties, this study proposes two new methods: (1) reformulating ODMTS-DA to a path-based single-level MIP (namely P-PATH) that can be solved to optimal in reasonable time and (2) five heuristic algorithms that rapidly find high quality solutions for large-scale instances. The computational benefits of both methods are demonstrated on two comprehensive case studies: the midsize transit system of the Ann Arbor – Ypsilanti region in Michigan and the large-scale transit system for the city of Atlanta. On the Michigan dataset, both methods can bring more than two orders of magnitude improvements compared with the exact algorithm. For Atlanta, the results show that P-PATH can solve large-scale ODMTS-DA instances optimally in a few hours or in a few days, and the heuristic algorithms can find high-quality solutions with even significantly less time (in a few minutes or in a few hours).

5 - Boosting Column Generation with Graph Neural Networks for Joint Rider Trip Planning and Crew Shift Scheduling

Tinghan Ye, Georgia Institute of Technology, Atlanta, GA, United States, Jiawei Lu, Wenbo Chen, Pascal Van Hentenryck

Optimizing service schedules is pivotal to the reliable, efficient, and inclusive on-demand mobility. This pressing challenge is further exacerbated by the increasing needs of an aging population, the oversubscription of existing services, and the lack of effective solution methods. This study addresses the intricacies of service scheduling, by jointly optimizing rider trip planning and crew scheduling for a complex dynamic mobility service. The resulting optimization problems are extremely challenging computationally for state-of-the-art methods.

To address this fundamental gap, this paper introduces the Joint Rider Trip Planning and Crew Shift Scheduling Problem (JRTPCSSP) and a novel solution method, called AGGNNI-CG (Attention and Gated GNN- Informed Column Generation), that hybridizes column generation and machine learning to obtain near-optimal solutions to the JRTPCSSP with the real-time constraints of the application. The key idea of the machine-learning component is to dramatically reduce the number of paths to explore in the pricing problem, accelerating the most time-consuming component of the column generation. The machine learning component is a graph neural network with an attention mechanism and a gated architecture, that is particularly suited to cater for the different input sizes coming from daily operations.

AGGNNI-CG has been applied to a challenging, real-world dataset from the Paratransit system of Chatham County in Georgia. It produces dramatic improvements compared to the baseline column generation approach, which typically cannot produce high-quality feasible solutions in reasonable time on both medium-sized and large-scale complex instances. AGGNNI-CG also produces significant improvements in service compared to the existing system.

SE78

Regency - 709

A Tutorial on Automated Decomposition Methods for Optimization

Invited Session

Computing Society

Chair: Matthew Galati, Amazon, Glen Mills, PA, United States

1 - Introduction to Decomposition Methods

Rob Pratt, SAS, Cary, NC, United States

Large-scale practical optimization problems typically have structure that can be exploited to reduce the solve time. In particular, the constraint matrices for mixed integer linear optimization problems often have block-angular form. Problems for which a small percentage of constraints do not appear in the blocks motivate Dantzig-Wolfe decomposition. On the other hand, problems for which a small percentage of variables do not appear in the blocks suggest Benders decomposition. This presentation provides an elementary introduction to both of these classical decomposition methods.

2 - Automated Decomposition Software

Matthew Galati, Amazon, Glen Mills, PA, United States

For problems with decomposable structure, Dantzig-Wolfe and Benders decomposition often outperform state-of-the-art implementations of the traditional branch-and-cut algorithm for mixed integer linear programming. Implementation of either decomposition algorithm can be cumbersome for specific applications, especially in cases that have multiple natural block decompositions, and many critical implementation details are not obvious from textbook descriptions. This presentation describes generic software frameworks that fully automate these two decomposition algorithms while still enabling the user to easily specify the block structure if desired.

3 - Benders Decomposition for the Critical Node Detection Problem

Laszlo Ladanyi, SAS, Lakeville, CT, United States

For an undirected graph, the critical node detection problem is to find a small subset of nodes whose removal yields the maximum disconnectivity of the resulting graph. This presentation motivates the problem, describes a mixed integer linear programming formulation, and illustrates the application of automated Benders decomposition to solve the problem efficiently.

4 - Decomposing Problem Data Instead of Problem Formulation: An Application to Assortment Optimization

Taghi Khaniyev, Bilkent University, Ankara, Turkey, Kaan Cakiroglu, Ali Ilhan Haliloglu, Elif Sena Isik, Elif Rana Yoner

Most decomposition approaches in linear integer programming such as Dantzig-Wolfe (DW) and Bender's become efficient alternatives to branch-and-cut based approaches by exploiting an inherent decomposable (i.e., bordered block diagonal) structure in the constraint matrices of the problems. For well-known problems such as facility location and vehicle routing, a decomposable structure is readily available from the problem formulation where problem can usually be decomposed into subproblems based on one of the index sets (e.g., customer set or potential locations of facilities). Many problems, however, do not have such a decomposable structure in their constraint matrices when problem data can take arbitrary values. In this work, we study the consider-then-choose assortment optimization problem where the generic formulation is not amenable to classical decomposition techniques. Many realistic problem instances, however, have such problem data (consideration matrix, in this case) that has a bordered block diagonal structure which lends itself to a new reformulation that can be exploited with DW decomposition. We provide an algorithm that can automatically detect such a decomposable structure in the problem data and consequently apply DW decomposition automatically. We conduct computational experiments on synthetic datasets to illustrate the efficient performance of the proposed reformulation.

Sunday, October 20, 5:25 PM - 6:15 PM

Summit - Ballroom 3

2024 INFORMS UPS George D. Smith Prize Reprise Presentation: University of South Carolina

Plenary/Keynote Session

Keynote

Chair: Anne Robinson, Robinson Insights, Ottawa, ON, Canada

1 - 2024 INFORMS UPS George D. Smith Prize Reprise Presentation: University of South Carolina

Sanjay Ahire, University of South Carolina, Columbia, SC, United States, Mark Ferguson

The Management Science Department in the Darla Moore School of Business at the University of South Carolina supports both undergraduate and graduate programs in operations and supply chain and business analytics as well as a Ph.D. in operations management. The department couples classroom and real-world, experiential learning to produce top-notch graduates. The Operations and Supply Chain (OSC) Center, through which students have the opportunity to earn an industry-certified Lean Six Sigma Green Belt, plays a central role in this real-world experience and has, over time, become a focal point for identifying, developing and placing well-trained talent in many leading Fortune 500 multinational firms. The OSC Center faculty have executed 360+ capstone consulting student projects in more than 45 partner firms yielding more than 330 million dollars in recurring benefits.

The department is also home to the Moore School Data Lab, which assists students with statistics and data analytics courses, and to the Center for Applied Business Analytics, which provides training, academic challenges and curriculum support to students and faculty. Graduate programs include the Masters of Science in Business Analytics and concentrations in Operations and Supply Chain in the school's various MBA programs. The department's executive education programs have trained over 5,000 officers in the U.S. Army on the use of advanced data analytics.

Doctoral graduates from the Management Science Department are currently working as faculty members at institutions such as Penn State University, The Ohio State University, Florida State University, University of Colorado, and University of Tennessee, among others.

Summit - Ballroom 2

Smarter Decisions Towards Preventing, Controlling, and Eliminating Infectious Diseases

Plenary/Keynote Session

Keynote

Chair: Maria Mayorga, North Carolina State University, Raleigh

1 - Smarter Decisions Towards Preventing, Controlling, and Eliminating Infectious Diseases

Pinar Keskinocak, ISyE Georgia Tech, Atlanta, GA, United States

Advances in science, medicine, and public health significantly improved our understanding of infectious diseases as well as their prevention and control; however, infectious diseases continue to have a devastating impact on the lives and livelihoods of millions people around the world. Global and regional changes in climate, human behavior, contact between human and wildlife reservoirs, and technology may further exacerbate the risk of infectious diseases. There are tremendous opportunities for analytics and operations research in combatting infectious diseases, especially in resource-constrained settings. We will discuss examples of how operations research and analytics can inform decision-making related to disease prevention and control, improve equity in access to resources and outcomes, and support efforts towards eradication.

Summit - Ballroom 1

Philip McCord Morse Lectureship

Plenary/Keynote Session

Keynote

Chair: Anahita Khojandi, University of Tennessee, Knoxville, TN, United States

1 - OR and Analytics in Support of Democracy

David Shmoys, Cornell University, Ithaca, NY, United States

Computational tools can now address (at scale) a wide range of decision-making elements of the democratic process. Perhaps the most well-studied of these is to construct (and evaluate) legislative districts for each state, a process dictated by the United States Constitution to recur every 10 years. Sixty years have passed since the first attempts at using integer programming .

This talk will briefly discuss several large-scale approaches made possible by both advances in IP and in computational power. Although these methods are effective, they provide opportunities to both achieve notions of fairness and to thwart such objectives. Innovative Markov Chain Monte Carlo (MCMC) methods have been introduced to provide a mathematical perspective to detect attempts at undermining fairness in the resulting districting plans.

One counterproposal advocated by several "good government" NGO's is to modify the current system to a "multi-member" one, where each (larger) district elects, for example, 3 representatives through a process of using ranked-choice votes with a mechanism called single-transferable voting. IP can again be used to analyze the impact of such a modified system, and demonstrates (at least at the first cut) that this change would both help to facilitate fair outcomes, and limit the power of those attempting to undermine such fairness. Finally, we will highlight the notion of a citizens' assembly and sketch how OR tools can be (and are, in practice) employed to deliver on the promise of delivering a fair democratic system by this alternative approach.

Monday, October 21, 8:00 AM - 9:15 AM

MA01

Summit - 320

Scheduling, Bike Sharing and Other Supply Chain Issues

Invited Session

Scheduling and Project Management

Chair: Zhi-Long Chen, University of Maryland, College Park, MD, 20742, United States

Co-Chair: Zhi-Long Chen, University of Maryland, College Park, MD, United States

1 - Online Integrated Production and Distribution Scheduling

Zhi-Long Chen, University of Maryland, College Park, MD, United States

A growing number of applications can be modeled as online integrated production and distribution scheduling (IPDS) problems. We provide several application examples, give a state-of-the-art review of the online IPDS research, and point out several directions for future research.

2 - Integrated Production and Transportation Scheduling Problem Under Nonlinear Cost Structures

Feng Li, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Julong Wang, Zhixue Liu

We consider an integrated production and transportation scheduling problem arising in several make-to-order settings, where customer orders have release times and pre-specified delivery time windows. These orders are first processed in a plant and, thereafter, delivered to their customer sites by transporters (e.g., freight trains, ships, or air flights) with fixed departure times and nonlinear transportation cost functions. If the processing of an order is completed but not immediately delivered, it will be stored temporarily, incurring inventory holding costs. The objective is finding an integrated schedule for production and transportation such that the total cost of inventory holding and transportation is minimized. In this study, we consider the following two cases: one where partial delivery is not allowed, and the other where partial delivery is allowed. For each case, we propose an exact algorithm to find optimal solutions for small-scale instances, and a heuristic algorithm to find near-optimal solutions for large-scale instances. For the randomly generated test instances, the computational results reveal that the exact algorithms significantly outperform a commercial optimization solver in terms of the computational times and the number of instances solved within a specified time limit, and the heuristic algorithms are capable of generating near-optimal solutions within a reasonable computational time. Finally, we incorporate considerations of sequence-dependent setup time into the problem and apply the approaches for designing the algorithms to address the more general problem.

3 - Supply Chain Configuration and Its Applications

Haitao Li, University of Missouri - St. Louis, Saint Louis, MO, United States

Simultaneously optimizing the system-wide safety stock placement and the modes/options to execute supply chain functions is known as the *supply chain configuration problem* (SCCP). With its general and versatile modeling framework, the SCCP and its variants have a plethora of applications in various industries. In this presentation, I will start with an introduction of the basic SCCP with its mathematical programming formulation and connection with a project scheduling problem. Extensions and applications are presented next, including the SCCP with resource constraints (SCCP-RC), a variant of the resource-constrained project scheduling problem (RCPSP), for make-to-order (MTO) manufacturing, and the food supply chain configuration problem (FSCCP) that considers food loss and quality deterioration to configure an end-to-end agricultural/food supply chain.

4 - A Static Mixed Bike Repositioning Problem with Both Man-Powered Bikes and Electric Bikes

Jia Cui, Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, Hong Kong, Wai Yuen Szeto

The proliferation of bike-sharing systems has significantly facilitated travel convenience and the sustainability of urban development. An introduction of electric bicycles (e-bikes) to existing bike-sharing systems can enhance the cycling experience by covering longer distances with less physical exertion. However, managing a bike-sharing system with both man-powered bikes and e-bikes presents a complex challenge, consisting of addressing the imbalance between bike supply, dock supply, and user demand. Despite this, no research has yet been conducted on this topic. This study presents a novel static bike repositioning problem that considers both man-powered bikes and e-bikes while allowing low-battery e-bikes to be recharged at charging stations. A mixed-integer linear programming model is proposed that aims to minimize the deviation from the target inventory of man-powered bikes and e-bikes (opposite to the deviation from the target inventory of vacant docks), the fixed cost of vehicles, the travel cost, and the loading cost within the time budget. Numerical experiments are conducted by a commercial solver on small network instances to show the effect of introducing (low-battery) e-bikes into existing bike-sharing systems. Efficient algorithms are also designed to provide near-optimal solutions to the problem under large instances. Computational experiments based on real-world data validate the performance of the algorithm and provide practical implications.

5 - Robotic Mobile Fulfillment Systems: Frameworks for Performance Analysis

Chelliah Sriskandarajah, Texas A&M University, College Station, TX, United States, Kerim Uygur Kizil, Jon Stauffer

Motivated by the growing demand for efficient order fulfillment in e-commerce, our study investigates three interconnected problems arising in managing robotic mobile fulfillment (RMF) systems: *(i)* allocation of SKU inventories into pods, *(ii)* pod selection for picking operations, and *(iii)* pod scheduling for picking operations. We deploy two distinct methodological approaches, modular and integrated, to explore how the level of integration in addressing the three problems influences overall system performance. System performance is defined by two metrics: *(i)* the total completion time (TCT) to fulfill all orders assigned to a specific interval and *(ii)* the number of required robots (NRR) to support seamless picking operations. By investigating pod scheduling as a standalone problem, we show it is NP-hard with two pickers [resp., one picker] under the TCT minimization [resp., NRR minimization] objective. Our findings suggest that the modular approach achieves an overall performance level comparable to the integrated one. In particular, we show that managers could divide and conquer the three problems by considering intuitive objective functions for each rather than having to rely on complex models that simultaneously address all aspects of RMF system operations.

MA02

Summit - 321

Data Science in Wind Energy

Invited Session

Quality, Statistics and Reliability

Chair: Eunshin Byon, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Yu Ding, Georgia Institute of Technology, Atlanta, GA, United States

1 - Explainable Importance Sampling for Reliability Analysis of Wind Turbines

Chenfei Li, University of Michigan, Ann Arbor, MI, United States, Jaeshin Park, Eunshin Byon

We propose a new explainable multivariate importance sampling method in stochastic simulation. Importance sampling easily faces the curse of dimensionality, as the input dimension grows. This study devises a new method that identifies crucial input variables to effectively formulate a nonparametric instrumental density and explains the importance of each variable. The wind turbine reliability case study demonstrates the method’s efficiency.

2 - Large Scale Data Handling for Wind Turbines

Ahmadreza Chokhachian, Georgia Institute of Technology, Atlanta, GA, United States, Matthais Katzfuss, Yu Ding

Wind power curve models connect environmental input with wind power output. Wind industry arranges their operational data in a 10 minutes interval, implying over 50,000 data pairs per turbine per year. When terrain information is to be modeled, multiple turbines on a wind farm need to be used to capture the spatial variation. A recent study using 66 wind turbines on a wind farm handles over 3.4 million data pairs. Handling such large datasets presents the computational challenge for using machine learning methods like the Gaussian process models. This talk presents our recent effort in speeding out the computation for handling large datasets for wind energy applications.

3 - A Hierarchical Optimization Framework to Integrate Turbine Failure Risks and Accessibility Forecasts to Offshore Wind Farm Operations, Maintenance, and Logistics

Ceyhan Sahin, Wayne State University, Detroit, MI, United States, Feng Ye, Ahmed Aziz Ezzat, Onur Kilic, Murat Yildirim

This paper introduces an innovative hierarchical optimization framework to optimize operations, maintenance, and logistics decisions in offshore wind farms. Operating at multiple levels, the framework leverages a monthly dynamic programming model at a higher tier to determine optimal vessel rental strategies, as a function of turbine statuses and future accessibility conditions. At a more granular level, a detailed two-stage model is developed to optimize operations on an hourly basis over a 30-day timeframe. Integrating various data-driven predictions such as availability forecasts, production levels across wind farm locations, and turbine failure risks, this framework offers a comprehensive approach to asset management in offshore wind farms. The proposed approach provides significant advantages over the benchmark policies in terms of both efficiency and reliability.

4 - Spatio-Temporal Offshore Wind Energy Forecasting over Space, Time, and Height

Ahmed Aziz Ezzat, Department of Industrial & Systems Engineering, Rutgers University, Piscataway, NJ, United States, Feng Ye, Xinxi Zhang, Michael Stein

To unlock access to stronger winds, the offshore wind industry is advancing with significantly larger and taller wind turbines. Existing spatio-temporal forecasting methods compress the three-dimensional reality of the wind field (space × time × height) into a simpler, two-dimensional “world view” (space × time) which completely overlooks the vertical dimension of the wind field. Instead, we propose a multi-output spatio-temporal model with a multivariate, nonstationary, and state-dependent kernel that is is capable of adequately modeling the complex dependencies in local wind speeds across multiple heights. Embedded within the statistical model, an advanced deep learning architecture fully exploits the wealth of exogenous information available to the forecaster (typically in the form of high-dimensional image streams) in order to learn and update the kernel parameters. Experiments using real-world data from the offshore wind energy areas in the Northeastern United States show that the wind speed and power forecasts from the proposed model are of considerably higher quality than those from prevalent time series, deep learning, and space-time methods.

MA03

Summit - 322

Cost-Efficient Computational Methods for Intelligent Systems

Invited Session

Quality, Statistics and Reliability

Chair: Mengfei Chen, The State University of New York at New Paltz, New Paltz, NY, United States

Co-Chair: Weihong Guo, Rutgers, The State University of New Jersey, Piscataway, NJ, United States

1 - Generative Model-Empowered Customized Metamaterial Design

Chenang Liu, Oklahoma State University, Stillwater, OK, United States, Ziyang Zhang, Yu Feng

This work presents a data-driven approach for designing customized metamaterial-structured shoe soles, tailored to individual plantar pressure using generative models. By integrating machine learning with finite element analysis, this work is expected to identify the optimal design based on user-specific biomechanical data. Specifically, our approach leverages plantar pressure measurements and preferences to adjust sole stiffness and cushioning, enhancing functionality and ergonomics while reducing development time and costs, thereby offering a scalable reverse design for personalized footwear.

2 - Robust Object Detection Using Unsupervised Multi-Domain Adaptation

Arpan Kusari, University of Michigan Transportation Research Institute, Ann Arbor, MI, United States, Hojun Son, Asma Almutairi, Venkata Sairam Polina, Jiahang Li

For learning based intelligent systems, the manual labeling process is labor intensive and costly. Transfer learning is an approach that enables cost-saving through use of other labeled datasets to train a neural network model and predict on our target dataset. However, it is challenging when confronted with significant distributional differences among datasets and noise in ground truth labels. To address these issues, an alternative which has been proposed in the domain adaptation realm is to learn from simulation and transfer onto a real environment, known as sim-to-real transfer. The simulation environment can produce infinite amount of data with large amount of control but lacks in variability which the real labelled datasets can provide. We devise a two-step domain adaptation approach towards object detection which takes into account the variability of multiple real datasets in learning a robust object detector model. We first train exhaustively on a photorealistic simulation environment, CARLA, altering the weather and daylight conditions along with the vehicle types seen on the road. Next, we simultaneously fine-tune on a number of real open-source object detection datasets in the automated vehicles domain. We provide some results on how the fine-tuning process can be applied, how the features of the background in the datasets can influence the foreground objects and the robustness of such object detection approaches as compared to learning from labeled examples. We show that such an approach can provide a very low-cost and powerful object detector.

3 - Plural: 3D Point Cloud Transfer Learning via Contrastive Learning with Augmentations

Michael Biehler, Georgia Institute of Technology, Atlanta, GA, United States, Jing Li, Jianjun Shi

Unlocking the power of 3D point cloud machine learning models can be a challenge due to the need for extensive labeled datasets, which presents a challenge when applying these models to new domains. Transfer learning can help overcome this challenge by utilizing data from related tasks to enhance model performance. However, traditional (2D) transfer learning methods struggle with 3D point cloud domain adaptation, due to differences in physical environments and sensor configurations. To address this issue, we propose PLURAL, a novel 3D point cloud transfer learning methodology based on contrastive learning with augmentations. Our approach is inspired by the notion that high-level shape features are more transferable than low-level geometry features. We propose a co-training architecture that includes separate 3D point cloud models with domain-specific parameters, as well as a module for learning domain-invariant features. Additionally, PLURAL extends the approach of contrastive instance alignment to 3D point cloud modeling by considering physics-informed hard sample mining. Our experiments on simulation and real-world datasets demonstrate that PLURAL outperforms state-of-the-art transfer learning methods by a significant margin, effectively reducing the domain gap.

4 - Heterogeneous Federated Learning via Generative Model-Aided Knowledge Distillation in the Edge

Chuanneng Sun, Rutgers University-New Brunswick, Edison, NJ, United States

Federated Learning (FL) has been popular recently as a framework for training Machine Learning (ML) models in a distributed and privacy-preserving manner. Traditional FL frameworks often struggle with model and statistical heterogeneity among participating clients, impacting learning performance and practicality. To overcome these fundamental limitations, we introduce Fed2KD+, a novel FL framework that leverages a set of tiny unified models and Conditional Variational Auto-Encoders (CVAEs) to enable FL training for heterogeneous models between network clients. Using forward and backward distillation processes, Fed2KD+ allows a seamless exchange of knowledge, mitigating data and heterogeneity problems of the model. Moreover, we propose a cosine similarity penalty in the loss function of CVAE+ to enhance the generalizability of CVAE for non-IID scenarios, improving the adaptability and efficiency of the framework. Additionally, our framework's design incorporates a codesign with Radio Access Network (RAN) architecture, reducing the fronthaul traffic volume and improving scalability. Extensive evaluations of one image and two IoT datasets demonstrate the superiority of Fed2KD+ in achieving higher accuracy and faster convergence compared to existing methods, including FedAvg, FedMD, and FedGen. Furthermore, we also performed hardware profiling on Raspberry Pi and NVIDIA Jetson Nano to quantify the additional resources required to train the unified and CVAE+ models.

5 - Estimation and Interpretation of the Effects of Process Variables in High-Temperature Superconductor tape manufacturing

Yisha Xiang, University of Houston, Houston, TX, United States, Yifan Huang, Zhigen Zhao

Improving the quality of the high-temperature superconductor (HTS) tapes is key to its commercialization. Identifying the significant process variables and interpreting their effects on the quality of the tape is important for the production. Stepwise linear regression is suggested to pinpoint the important predictors. Principal Component Analysis was employed to determine a few uncorrelated linear combinations of the process variables that account for most of their variation. There are some major drawbacks to these existing approaches, including the linearity assumption and underutilization of the response variable. To address these issues, we will develop a model-free screening and estimation method to identify critical process variables and find a sufficient low-dimensional representation for ultra-high dimensional data to assess the quality of the tapes.

MA04

Summit - 323

System Dynamics and Sensor-Driven Data Analytics in Complex System Reliability

Invited Session

Quality, Statistics and Reliability

Chair: Phat Huynh, North Carolina A&T State University, Greensboro, NC, 27411, United States

Co-Chair: Trung Le, University of South Florida, Tampa, FL, United States

1 - Physics-Informed Multi-Scale System Reliability Modeling and Digital Twins in Multi-Layered Smart Grid Networks

Phat Huynh, North Carolina A&T State University, Greensboro, NC, United States, Trung Le

Smart grid systems, characterized by their intricate multi-layered architectures and susceptibility to cascading failures, demand a novel approach to reliability modeling that incorporates both real-time data and comprehensive system dynamics. This study bridges significant gaps in existing methodologies by introducing a sophisticated, physics-informed multi-scale modeling framework that leverages the synergy between stochastic system dynamics models and digital twin technology. In our methodology, we first formulate the problem by identifying the stochastic nature of grid failures and the cascading effects that can result from interdependencies across different layers of the grid. We develop a set of differential equations that describe the physics of electrical flow and the conditions leading to system instabilities, informed by historical PMU data and failure logs. Using this foundation, we construct a physics-informed stochastic system dynamics model that simulates the interaction between different grid components and their failure mechanisms. This model quantitatively predicts the likelihood and impact of cascading failures by simulating various failure scenarios under different conditions. To dynamically integrate these simulations with real-time operational data, we implement a digital twin of the smart grid. This digital twin serves as a high-fidelity surrogate model, continuously updated with real-time PMU data. It not only simulates current grid conditions but also forecasts future states under potential stress scenarios. This study not only advances the field of system reliability but also sets a robust framework for the application of physics-informed modeling and digital twins in complex, interconnected energy systems.

2 - Adaptive Multi-scale Threshold Modeling for Intermittency Dynamics by Multivariate Koopman Spectral Analysis: A Case Study in Pathophysiological Processes of Obstructive Sleep Apnea

Viet Nguyen Ca Hoang, University of South Florida, Tampa, FL, United States, Ethan Kessler, Phat Huynh, Trung Le

This study introduces a comprehensive framework to address the intermittent and complex nature of obstructive sleep apnea (OSA). We propose a novel approach that integrates multivariate time series analysis with adaptive multi-scale dynamics to improve detection and prediction of intermittent physiological events characteristic of OSA. Our methodology begins with state space reconstruction using Takens' embedding theorem on high-dimensional physiological data to capture intrinsic system dynamics, including temporal interactions. We then implement multi-scale entropy analysis to quantify signal complexity across temporal scales, which is crucial for identifying transient dynamics and distinguishing between stable and unstable respiratory phases indicative of OSA episodes. Analyzing signal complexity at multiple scales provides a more nuanced understanding of OSA's intermittent nature, enabling precise identification of critical transitions in breathing patterns. Central to our framework is an advanced adaptive threshold model based on autoregressive techniques. This model dynamically adjusts detection thresholds in real-time, tailored to the multivariate and multi-scale physiological data, through an algorithm that forecasts future states of intermittency using linear and non-linear autoregressive models. This adaptive approach accommodates patient-specific variability and the non-stationarity nature of OSA-related signals. Our integrated approach, validated on a diverse dataset of OSA patients, demonstrated a 40% increase in detection sensitivity over traditional models and adaptability to evolving OSA dynamics. This research advances physiological monitoring and diagnostic analytics by providing a robust, scalable, and adaptive tool for managing OSA. It sets new standards in predictive modeling of complex health conditions, paving the way for innovations in medical technology and patient care.

3 - Longitudinal Survival Analysis in Traumatic Brain Injury Cohorts with Diverse Obstructive Sleep Apnea Endophenotypes

Tien Truong, University of South Florida, Tampa, FL, United States, Phat Huynh, Trung Le

Traumatic Brain Injury (TBI) and Obstructive Sleep Apnea (OSA) are significant public health concerns due to their high incidence and complex interdependencies which impact patient rehabilitation and outcomes. This study identifies OSA-TBI endophenotypes and analyzes their survival patterns and readmission factors to improve patient management and outcomes. Our methodology integrates innovative steps, utilizing advanced clustering algorithms to analyze clinical, physiological, and behavioral data, identifying unique patient endophenotypes with shared characteristics, and enhancing the precision and applicability of survival models for patients with combined OSA-TBI. Subsequently, employing state-space models, we develop a dynamic survival analysis framework that incorporates time-varying health states and continuous health monitoring data, allowing for real-time adjustments based on patient condition changes. We integrate hospital readmission data as a dynamic covariate within the survival model, examining its influence on patient outcomes across endophenotypes, and validate the models using advanced statistical techniques to ensure robustness and reliability. Preliminary results show significant variations in survival rates and readmission causes among OSA-TBI endophenotypes, with dynamic survival models outperforming traditional static models in predicting short-term and long-term patient outcomes. Integrating cluster analysis with dynamic modeling provides a comprehensive view of patient trajectories and risk factors, enabling tailored interventions, while the novel use of state-space models for real-time data integration advances predictive analytics for complex health conditions. This study advances survival analysis in complex clinical scenarios, providing tools to understand and predict TBI and OSA co-morbidity trajectories, and offering endophenotype-based insights to improve personalized care and reduce unnecessary readmissions.

4 - Tests for Homogeneity of Component Lifetime Distributions Based on System Lifetime Data

Tony Ng, Bentley University, Waltham, MA, United States, Jingjing Qu

In system reliability engineering, systems are made up of different components, and these systems can be complex. For various purposes, engineers and researchers are often interested in the lifetime distribution of the system as well as the lifetime distribution of the components that make up the system. In many cases, the lifetimes of an n-component coherent system can be observed, but not the lifetimes of the components. In recent years, parametric and nonparametric inference for the lifetime distribution of components based on system lifetime lifetimes has been developed. In this talk, we discuss the problem of testing the homogeneity of component lifetime distributions based on system lifetime data with known system signatures. Existing test procedures for the homogeneity of component lifetime distributions based on complete system lifetime data are reviewed. Then, several nonparametric testing statistics based on the empirical likelihood method are proposed for testing the homogeneity of two or more component lifetime distributions. Both complete and Type-II censored system lifetime data will be considered. The performance of the proposed empirical likelihood ratio tests is compared with other parametric and nonparametric tests in the literature. Finally, some concluding remarks and possible future research directions are provided.

MA05

Summit - 324

Diversity, Equity and Inclusion Special Issue of POM: Impacts and Implications for Scholarship

Panel Session

Diversity, Equity, and Inclusion

Co-Chair: Michael Johnson, University of Massachusetts Boston, Boston, MA, 02125, United States

1 - Moderator Panelist

Michael Johnson, University of Massachusetts Boston, Boston, MA, United States

2 - Panelist

Wiljeana Glover, Babson University, Fayetteville, MA, United States

Dr. Wiljeana Glover was an associate editor for the Production and Operations Management special issue on diversity, equity and inclusion.

3 - Panelist

Charles Corbett, UCLA Anderson School of Management, Los Angeles, CA, United States

Charles Corbett was co-editor of the Production and Operations Management special issue on diversity, equity and inclusion.

4 - Panelist

Sriram Narayanan, Michigan State University, East Lansing, MI, United States

Sriram Narayanan was co-editor of the Production and Operations Management special issue on diversity, equity and inclusion.

5 - Panelist

Subodha Kumar, Fox School of Business, Temple University, Philadelphia, PA, United States

Subodha Kumar is deputy editor of Production and Operations Management.

6 - Panelist

Margret Bjarnadottir, University of Maryland at College Park Robert H Smith School of Business, College Park, MD, United States

MA06

Summit - 325

Location Models

Invited Session

Location Analysis

Chair: Zvi Drezner, California State University Fullerton, Fullerton, CA, United States

1 - Optimizing Shape and Compactness

Alan Murray, University of California at Santa Barbara, Santa Barbara, CA, United States

Geographic shape has long been an intriguing feature of observed and defined facets of an area or region. Compactness reflects a critical element of shape with important practical and policy implications for urban form, trade and service provision, political representation and physical phenomena. While there has been much study of compactness and a wealth of measures and metrics derived to reflect nuances of geographic form, questions remain about their ability to characterize shape in a meaningful manner. This paper reviews representative measures compactness, examining both theoretical and practical issues as well as utilization in optimization.

2 - Locating Obnoxious Facilities With Variable Capacities

Zvi Drezner, California State University Fullerton, Fullerton, CA, United States

The problem of locating facilities that have a negative impact on surrounding communities was initiated about 50 years ago, and intensively analyzed in the literature. In the models proposed in the literature, it is assumed that all facilities have the same capacity (size). We propose to construct facilities of different capacities. Facilities located in open areas are far from communities and can be larger without impacting the most affected community. It allows facilities close to communities to be smaller without changing the total capacity and significantly reduces the negative impact on the most affected community. Such an important extension leads to a more complicated formulation that we were able to simplify. The simplified formulations resulted in much better solutions in shorter run times than the more complicated formulations. The specially designed algorithms proposed in this paper find much better configurations than those found for facilities that have the same capacities. In extensive computational experiments we found in some cases that the negative impact on the most affected community is reduced by more than 50% when facilities of different capacities are established.

3 - On Fairness-Efficiency Trade-offs in Service Network Design

mina torabi milani, University of Waterloo, Waterloo, ON, Canada, Hossein Abouee Mehrizi, Hamid Arzani, Saeed Ghadimi

In urban regions, budget-constrained service providers often face critical trade-offs between efficiency and fair access to service for individuals. In this paper, we study a service network design and resource allocation problem, where the service provider should serve the demand of customers residing along a line segment by deciding on the number of service facilities to locate, their locations, and the resources allocated to each, while adhering to budget constraints. Customers seeking service are self-interested, i.e., each aims to minimize their disutility, expressed as the sum of their travel and wait costs. We focus on two service providers: the efficient (i.e., utilitarian) provider, whose goal is to minimize the expected total disutility of customers, and the fair (i.e., egalitarian) provider, who aims to minimize the maximum disutility between the customers. We demonstrate that when the wait cost dominates the travel cost, the optimal fair policy closely follows the optimal efficient policy. In the regions where two policies disagree, we analyze the relative efficiency loss, namely the price of fairness (POF), associated with the fair provider. We derive analytical bounds on the POF for a broad class of demand distributions. We also examine how the spatial distribution of the population affects the POF, showing that a more dispersed distribution generally results in a lower POF.

Moreover, we observe that a population concentration near the urban center reduces the POF under some conditions. Conversely, in some cases, the POF decreases as the population concentration moves away from the center.

4 - Heterogeneous Directional Sensor Coverage Problem

Enbo Zhou, University of California, Santa Barbara, Santa Barbara, CA, United States, Alan Murray

Sensor network placement is critical for addressing a range of urban issues and concerns, including security surveillance, ecological observation, traffic monitoring, internet of things, and others. The appropriate deployment of a sensor network can increase system efficiency and reduce operation costs. Most existing studies assume identical sensors with isotropic coverage. However, sensors may be heterogeneous in practice because of different purchase time, budget limit and application requirements. Heterogenous sensors may have different specifications in viewing radius, field of view, operational spectrum and battery longevity, among others. It remains a question how to strategically select the sensor type, location, and orientation for a sensor network. To address these issues, a spatial optimization model is developed to simultaneously consider sensor type, location, orientation, and occlusion. Visibility analysis by sensor type is an important spatial component. Further, a finite dominating set can be derived for each sensor orientation, enabling linear integer program to be formulated. The proposed model is applied to a surveillance camera deployment case study on the UC Santa Barbara campus, demonstrating its effectiveness. The results highlight a holistic solution approach for sensor network planning, security surveillance, and emergency response.

5 - A Machine Learning Enhanced Column Generation in Location-Allocation with Equity Consideration for Efficient Vaccine Distribution

Kuangying Li, North Carolina State University, Raleigh, NC, United States, Hiruni Niwunhella, Leila Hajibabai, Ali Hajbabaie

This study presents a maximal covering location-allocation problem, formulated as mixed-integer linear model, that aims to minimize the total cost of vaccine distribution and maximize vaccine allocation to population blocks including equity constraints for age, race, and gender groups. A modified Voronoi diagram technique embedded in a Lagrangian relaxation framework is employed to solve the problem. A set of empirical case studies in Pennsylvania are conducted based on the real-world data obtained from the Centers for Disease Control and Prevention (CDC) and health department websites. The numerical experiments indicate that the proposed model can solve the problem effectively and outperform those of a benchmark column generation technique. While these methods effectively tackle the distribution problem, they both experience significant CPU times due to complex decision variables and extensive data. Therefore, this study proposes a machine learning-enhanced column generation method to improve computational efficiency while preserving solution quality, effectively reducing runtime and maintaining an acceptable optimality gap in vaccine distribution. The preliminary results indicate that the proposed approach can effectively solve the problem and outperform the solutions obtained from our prior work. The enhanced performance results from using predictive analytics within the machine learning process, which significantly improves the model’s capacity to manage the complexities associated with integer decision variables.

MA07

Summit - 327

Optimization and Decision Models

Flash Session

Chair: Alex Savachkin, UNIVERSITY OF SOUTH FLORIDA, Tampa, FL, United States

1 - Forecasting Intermittent & Declining Demand in the End-of-Life Phase

Robyn Goldsmith, Lancaster University, Lancaster, United Kingdom, Anna-Lena Sachs

Across a number of sectors, after-sales services constitute a billion-dollar business. Final purchases are high stake decisions of particular importance for the aftermarket business. They are often made towards the end of the spare part life cycle and need to balance shortages and over-ordering so as to sustain demand for the rest of a part’s life. In the automotive industry, among others, this period can extend for as long as ten years.

Forecasting in order to make effective final purchase decisions is made difficult by the nature of the demand. A large majority of spare part inventories consist of items with intermittent demand structures, where in some periods no demand is observed at all. As a result, standard forecasting methods provide inaccurate forecasts for intermittent items. As final purchase decisions are frequently made to sustain demand several years into the future, modelling procedures should also be designed to include the decay in demand as time passes. A forecasting model that accounts for aspects of both decline and the intermittency of demand are the subject of this talk.

2 - Kondo

Ankit Jain, Amazon, Bellevue, WA, United States

Amazon's inventory is the result of complex stochastic decision-making processes that consider various economic, supply chain, and operational factors. These decisions are aimed at delighting customers by providing them with a wide range of selection at affordable prices with speedy delivery and lower cost to serve. However, until recently, there has not been a direct way to link the amount of inventory held with customer experience (delivery speed) and cost to serve the customers and audit these decision-making processes in a scalable way.

In our research, we focus on outbound shipping cost which is one of the main components of cost to serve. One of the key findings of our research is that inventory that optimizes for customer experience already captures most of the outbound shipping cost savings opportunity. Hence, from an inventory perspective, optimizing for customer experience is similar to optimizing for outbound shipping cost. Our approach is unique in that we use machine learning in combination with observational data patterns to uncover the relationship between inventory outcomes, customer experience and outbound shipping cost in a scalable way.

3 - A constraint programming approach to resource allocation for tour group travel planning

Xinran Ma, Mobi Systems, Somerville , MA, United States, Edmund Williams, Cheng Fang, Jacob Broida, Peng Yu

Tour and travel planning is one of the leading industries, accounting for approximately 10% of the world's GDP. The problem of tour group planning is particularly challenging for providers. Given an inventory of hotel rooms and restaurant places, the provider is required to allocate room and meal reservations to groups of travellers, and construct itineraries for places of interest consistent with timing constraints and travel routing times.

We decomposed the problem into two subproblems, one of allocating the resources, and one of itinerary construction given the chosen hotel and restaurants. At the top level, we map the resource allocation problem to one of bin packing. We then check for the feasibility of the resource allocation with respect to the travel time constraints, by formulating the itinerary creation problem as a vehicle routing problem with time windows. We iteratively generate resource allocations, guided by results of the bottom level, and use local search to improve solutions.

We will provide a detailed description of the encoding, as well as the solution method. We will provide an evaluation of the approach on datasets of hotels, restaurants, and places of interest inspired by real-world tour group travel destinations.

4 - Linear programming optimization model for repetitive prefabricated construction projects considering renewable resource categories

Dingfeng Yang, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

Prefabricated construction refers to buildings where prefabricated components are produced in factories and then transported to construction sites for sequential assembly by lifting.The construction of multi-building, multi-story projects is a repetitive project. In traditional Resource-Constrained Project Scheduling Problems in repetitive activities projects(RCPSP-RA), renewable resources typically refer to those available for all tasks and units within the project. However, in prefabricated construction projects, due to resource complexity, their working scope needs consideration. For example, tower cranes are limited to tasks within their jib's reach, climbing frames and hoists operate within a single building. Therefore, this paper categorized renewable resources into four types: tower crane resources, building resources, crew resources, and global resources. Additionally, a new precedence relationship—spatial precedence relationship—was proposed to describe the relationship between different tasks on adjacent floors.Subsequently, a bi-objective linear programming model was constructed to schedule the start times of each activity, aiming to minimize both the project duration and the interruption time of crews, without violating feasibility constraints.The model was developed in three main phases: 1) identifying input data, 2) optimizing the constraints schedule of repetitive units by determining decision variables, formulating objective functions and constraints, and 3) visualizing output results.Finally,one case study of prefabricated building was analyzed to evaluate the model performance.Moreover,The model provided practitioners with a clear framework and could also serve as a guide for newcomers to understand and extend scheduling problems in prefabricated repetitive projects.

5 - Leveraging Smart Contracts to Improve Construction Supply Chain Agility During Emergencies

Zhujun Wang, School of Management, Xi’an Jiaotong University, Xi'an, Chile, Qin Su , Miroslaw Skibniewski

Construction projects belong to capital-intensive industries, and the structure of the construction supply chain is complex, making it susceptible to external factors. Typically, multiple stakeholders are involved in the construction supply chain, each with unique and sometimes overlapping business interests. Consequently, the supply chain often struggles to respond swiftly to emergencies, potentially leading to significant economic losses.

This paper introduces a smart contract prototype designed to enhance the agility of the construction supply chain during emergencies. This prototype, deployed on a blockchain, effectively links all stakeholders within the supply chain into a unified community. This integration significantly improves the efficiency and trust level of information transmission throughout the chain. Additionally, the prototype integrates with group decision-making models external to the blockchain platform, enabling all stakeholders in the supply chain to participate in decision-making. This integration fosters the alignment of stakeholder interests and facilitates the formation of well-consented decisions rapidly to manage emergencies effectively. The smart contract prototype can automatically execute decision results and log all transaction information for stakeholders to review. This paper employs Ethereum to simulate the prototype, demonstrating its practical viability.

This research contributes a novel smart contract-based solution to managing emergencies within construction supply chains, thus improving their collaborative efficiency and operational agility. Future studies are encouraged to explore the design of information transparency levels within the smart contract prototype to enhance its real-world applicability.

6 - Modeling Antiviral Resistance for Mitigating Pandemic Influenza Outbreaks

Alex Savachkin, University of South Florida, Tampa, FL, United States, Sandro Paz

The response to the next pandemic influenza will likely include extensive use of antiviral drugs, which will create significant selective pressure for the emergence of antiviral resistant strains. The existing literature has inadequate decision models to support mitigation of a pandemic influenza born by an antiviral resistant strain. We present a large-scale simulation optimization model for developing dynamic antiviral strategies including treatment of symptomatic cases and chemoprophylaxis of pre- and post-exposure cases. The model considers an oseltamivir-sensitive strain and a resistant strain with low/high fitness cost, induced by the use of the several antiviral measures. The model is tested on a hypothetical regional outbreak, involving more than one million people.

MA08

Summit - 328

Transportation Analytics in Military and Security Applications

Invited Session

Military and Security

Chair: Jesse Pietz, MIT Lincoln Laboratory, Lexington, MA, United States

Co-Chair: Coralys Colon-Morales, MIT Lincoln Laboratory, Woburn, MA, 01801, United States

1 - Aircrew Scheduling Optimization

Coralys Colon-Morales, MIT Lincoln Laboratory, Woburn, MA, United States

Personnel scheduling is a well-known challenge across a wide variety of domains. Our work on Puckboard demonstrates a real-world application of personnel scheduling to flight crews in the U.S. Air Force. We utilize both linear & constraint optimization techniques to recommend air-crew based on factors such as legality, currency, and training history. Additionally, we encode a number of domain specific constraints such as flying time limits and instructor waivers. Using real personnel, mission, & training data, we provide optimal crew sets to the scheduler for review along with justification statements for explainability. As the application is used by schedulers, their assignment decisions are recorded alongside the recommendations of our algorithms, allowing for both refinement of the constraints and user workflow adjustments to ensure optimal and robust schedules.

2 - An Optimally Controlled Spatially-directed SIRS Epidemiology Model Driven by Multifractional Brownian Noise

Steven Kusiak, MIT Lincoln Laboratory, Lexington, MA, United States

The COVID-19 pandemic emphasized the important need for robust forecasting and advanced modeling frameworks that better describe the spread of highly contagious diseases in complex environments and under considerable uncertainty. Equally important, or more so, the pandemic revealed the need for tools that can leverage such models and determine the optimally balanced use of non-pharmaceutical interventions that can assist in controlling the spread of disease while simultaneously allowing for maximal in-class school engagement, in-office work, and worldwide production and delivery of critical goods and services. We present a novel analytical framework that casts generalized stochastic Susceptible-Infected-Recovered (SIR)-like compartmental disease models driven by time-dependent fractional Brownian noise into deterministic time evolutionary forward and backward partial differential equations whose solutions may be controlled. The model has allowed us to create an optimal control framework that provides the best tradeoff between an acceptable and controlled degree of disease spread with the largest amount of population interaction and migration between population centers. The model sheds much insight into the dynamic nature of diseases such as COVID-19 and provides a computational tool that determines optimal actionable policies in support of pandemic management under considerable uncertainty.

3 - The Backbone of Global Reach: Air Refueling Scheduling Algorithms for Air Campaign Modeling

Jesse Pietz, MIT Lincoln Laboratory, Lexington, MA, United States, Coralys Colon-Morales

In order to estimate the demand for refueling aircraft, U.S. Air Force planners must model air campaign scenarios. In each scenario refueling aircraft are scheduled to meet the fuel needs of receiver aircraft. This modeling effort leads to a difficult vehicle routing optimization problem that must be solved quickly as planners iterate through a large collection of scenarios. This presentation highlights key features of this refueling aircraft routing problem and describes a benchmark greedy algorithm that is used to compute high-quality solutions. In order to improve solution quality and decrease runtime, five algorithm variants are considered. Algorithms are tested against a library of 4,000 randomly-generated scenarios, and numerical results are presented to illustrate relative algorithm performance.

4 - Vehicle Routing Problem Approach for Improving Fuel Delivery Scheduling to Austere Sites

Brian Lemay, United States Air Force Academy, Colorado Springs, CO, United States, Jared Stewart, Kate Leonard, Gentry Lamb, Jack Heller, Sophie Allen

The 96th Logistics Readiness Squadron at Eglin Air Force Base, Florida conducts an average of 55 million gallons of annual refueling operations across the 464,000-acre Eglin Reservation. Currently, ground refueling operations plans generate suboptimal refueling routes, negatively affecting fuel consumption and operational costs. To improve the efficiency of the refueling operations, we created a data-driven scheduling tool inspired by the Vehicle Routing Problem. The study focuses on 20 sites across the reservation. When applying our scheduling tool to historical refueling operations, we find there are opportunities to significantly reduce mileage on refueling vehicles. The study aims to validate the effectiveness of the proposed tool by comparing suggested routes with current operations and offering actionable recommendations for future fuel distribution logistics. The analysis revealed an average of 6,650 miles saved per year. The scheduling tool's flexibility allows for adapted use on other installations and different operations.

MA09

Summit - 329

Social Good Applications

Invited Session

Auctions and Market Design

Chair: Elisabeth Paulson, Harvard Business School, Somerville, MA, United States

1 - Digitized Indian Micro-Grocery Transactions Reveal that Grain Subsidies Reduce Junk Food Buying by Low-Income Shoppers

Kamalini Ramdas, London Business School, London, United Kingdom, Ali Aouad, Alp Sungu

What foods do people living on under :$2 a day buy? How do food subsidies affect their food shopping? The digitization of 39 micro-retail groceries and food vendors in a Mumbai settlement via installing point-of-sale scanners provides novel and objective data on individual shoppers' grocery transactions. Our data reveal that individuals residing in households with children {purchase} significantly more junk foods, and our experimental evidence suggests that staple food subsidies cause demand substitution towards a healthier diet. Based on over half a million food transactions made by 23,717 consumers, snacks, sugar and soft drinks represent 14.5 percent of food spending in our grocery stores, while rice and wheat combined are 15.6 percent. We randomized 1,255 settlement residents into a weekly rice-and-wheat subsidy disbursed via a government-like subsidy store – or none, for six weeks. Treated subjects substituted away from packaged snacks, sugar and soft drinks spending by 20 to 30 percent more than untreated subjects and increased spices and accompaniments spending by over 30 percent more, with strongest effects for those in households with children. Overall food spending, calories, and nutrients from purchased foods remained unchanged, indicating that the subsidy likely increased the total demand for food and total consumed nutrients.

2 - Load Estimation, Optimal Sizing, and Simulation of Micromobility-to-Microgrid Systems for Health Facility Electrification

Rebecca Alcock, University of Wisconsin–Madison, Madison, WI, United States, Justin Boutilier

Persisting energy poverty in healthcare facilities endangers patients and providers alike by limiting available services and creating poor working conditions. For clinics not yet connected to a traditional power grid, distributed clean energy solutions are the timeliest and most climate-resilient option. However, without accessible and relevant renewable energy design tools, these facilities must wait for grid connection or turn to carbon-based decentralized alternatives. Motivated by the urgent and widening nature of the energy-health crisis, we are developing a toolkit to optimally design a solar, storage, and e-vehicle network for never-before-electrified health facilities. The toolkit comprises three components: 1) predictive analytics for parameter generation, 2) prescriptive analytics to size the energy system, and 3) simulation to evaluate long-run performance under changing climate and healthcare conditions. We explore the results of the toolkit for a case study in rural Colombia.

3 - Estimating Heterogeneous Treatment Effects in Panel Data

Emily Zhang, Massachusetts Institute of Technology, Cambridge, MA, United States, Retsef Levi, Elisabeth Paulson, Georgia Perakis

This work addresses a core problem in causal inference: estimating heterogeneous treatment effects with complex intervention patterns using panel data. Motivated by our ongoing collaboration with the Massachusetts Department of Transitional Assistance (DTA), our methodology also extends to broader applications. Specifically, we make the following contributions. 1) We develop a novel causal inference method for this complex setting. 2) Our theoretical results establish the convergence of the resulting estimates. 3) We perform computation experiments on synthetic data, demonstrating that our method achieves superior accuracy compared to current state-of-the-art approaches. These analytical and computational findings reinforce our confidence in the method's reliability for real-world applications.

4 - Notifications and Social Media Activity: Evidence from a Field Experiment

Ananya Sen, Carnegie Mellon University, Pittsburgh, PA, United States, Alp Sungu, Saharsh Agarwal

Policymakers and social media platform managers coming to terms with digital addiction and its adverse societal consequences. Legislation (eg. SMART) being put forward that suggest design based interventions (eg no infinite scroll) to curb problematic social media use. We examine the impact of regulating people's access to social media notifications on their digital activity and well being. We focus on the behavior of teenagers, which has been the focus of policy conversations.

Notifications offer `intermittent variable rewards', and thus have all the necessary ingredients to trigger fear of missing out (FOMO). We implement a field experiment to analyze the impact of regulating notifications on real digital activity, including cross-app substitution. We recruit a large sample of teenagers in India that allows us to provide a rich understanding of the impact of such a policy recommendation. We also aim to benchmark these effects against session-specific time limits that aim to limit infinite scrolling. Individuals are heterogenous in their self-control and self-regulation abilities, and we also examine how these differences influence their susceptibility to the addictive effects of social media notifications. Not all notifications are the same, and we examine how the different types of social media notifications (e.g., likes, comments, direct messages) contribute to impulsive social media use. Finally, we look at the effect of treatment reversal by analyzing what happens to social media activity when the regulation on notification delivery is lifted. Answers to these questions will help quantify existing welfare losses driven by social media notifications.

MA10

Summit - 330

Combinatorial Auctions / LLMs

Invited Session

Auctions and Market Design

Chair: Benjamin Lubin, Boston University, Boston, MA, 02215, United States

1 - Multidimensional Mechanism Design with Side Information

Siddharth Prasad, Carnegie Mellon University, Pittsburgh, PA, United States, Nina Balcan, Tuomas Sandholm

We develop a general tunable framework for multidimensional mechanism design with side information to boost revenue while preserving efficiency and incentives. Side information can come from a variety of sources— examples include advice from a domain expert, predictions from a machine-learning model trained on historical agent data, or even the mechanism designer's own gut instinct—and in practice such sources are abundant. We design a tunable mechanism that integrates input side information with an improvement of the classical VCG mechanism. The welfare, revenue, and incentive properties of our mechanism are characterized by a number of novel constructions we introduce based on the notion of a weakest competitor, which is an agent that has the smallest impact on welfare. When side information is of reasonably high quality and our mechanism is tuned well, it extracts welfare and revenue competitive with the total social surplus, and its performance degrades continuously and gradually as both the quality of the side information and the mechanism tuning worsens.

Time permitting, I will discuss computational challenges that arise when applying our approach to combinatorial auctions.

2 - Ad Auctions for LLMs via Retrieval Augmented Generation

Sébastien Lahaie, Google Research, New York, NY, United States, Suho Shin, Keivan Rezaei, MohammadTaghi Hajiaghayi

In the field of computational advertising, the integration of ads into the outputs of large language models (LLMs) presents an opportunity to support these services without compromising content integrity. This paper introduces novel auction mechanisms for ad allocation and pricing within the textual outputs of LLMs, leveraging retrieval-augmented generation (RAG). We propose a segment auction where an ad is probabilistically retrieved for each discourse segment (paragraph, section, or entire output) according to its bid and relevance, following the RAG framework, and priced according to competing bids. We provide a theoretical analysis of the efficiency, revenue, and incentive properties of various pricing rules, including a characterization of the incentive-compatible pricing scheme. An empirical evaluation validates the feasibility and effectiveness of our approach.

3 - Pricing Valid Cuts for Walrasian and Price-Match Equilibria

Robert Day, University of Connecticut, Storrs, CT, United States, Benjamin Lubin

In a combinatorial auction, there is often no vector of prices for the items being auctioned that constitutes a Walrasian equilibrium, in which all bidders prefer their allocated bundle and excess supply items are priced at zero. We use valid inequalities (cuts) of the binary integer program for winner determination as "artificial items" that can be interpreted intuitively and priced to generate an "Artificial Walrasian Equilibrium." We show that the lack of an integer programming gap is not a sufficient stopping criterion for cut generation to arrive at a "Price-Match Equilibrium" and provide examples where this refinement differs from natural Walrasian equilibrium. We prove the existence and characterize "Artificial Walrasian Equilibria" and prescribe "Minimally Artificial Price-Match" payments as a desirable equilibrium selection.

4 - Truthful Aggregation of LLMs with an Application to Online Advertising

Ermis Nikiforos Soumalias, University of Zurich, Zurich, Switzerland, Michael Curry, Sven Seuken

We address the challenge of aggregating the preferences of multiple agents over LLM-generated replies to user queries, where agents might modify or exaggerate their preferences. New agents may participate for each new query, making fine-tuning LLMs on these preferences impractical. To overcome these challenges, we propose an auction mechanism that operates without fine-tuning or access to model weights. This mechanism is designed to provably converge to the output of an optimally fine-tuned model as computational resources are increased.The mechanism can also incorporate contextual information about the agents when available, which significantly accelerates its convergence. A well-designed payment rule ensures that truthful reporting is the optimal strategy for all agents, while also promoting an equity property by aligning each agent’s utility with her contribution to social welfare, an essential feature for the mechanism's long-term viability. While our approach can be applied whenever monetary transactions are permissible, our flagship application is in online advertising. In this context, advertisers try to steer LLM-generated responses towards their brand interests, while the platform aims to maximize advertiser value and ensure user satisfaction. Experimental results confirm that our mechanism not only converges efficiently to the optimally fine-tuned LLM but also significantly boosts advertiser value and platform revenue, all with minimal computational overhead.

MA12

Summit - 332

Logistics Optimization: Routing and Delivery

Flash Session

Chair: Dawei Chen, Eindhoven University of Technology, Eindhoven, 5612AH

1 - A Deep Reinforcement Learning Approach for the Dynamic Travelling Salesman Problem with Time-Dependent and Stochastic Travel Times

Dawei Chen, Eindhoven University of Technology, Eindhoven, Netherlands, Christina Imdahl, David Lai, Tom Van Woensel

We introduce a novel approach using deep reinforcement learning to tackle the Dynamic Traveling Salesman Problem with Time-Dependent and Stochastic travel times (DTSP-TDS). The main goal is to dynamically plan the route with the shortest tour duration and visit all customers while considering the stochastic nature of travel times. We employ a reinforcement learning approach to dynamically address the stochastic travel times to observe changing states and make decisions accordingly. Our reinforcement learning approach incorporates a Dynamic Graph Temporal Attention model (DGTA) that dynamically extracts information from the stochastic environment. We have tested the performance of our proposed approach on the simulation. Our approach can quickly provide high-quality solutions for all datasets using limited computing resources. The efficiency testing on different models demonstrates that new components in the DGTA model, such as the selection mechanism, dynamic attention component, and temporal pointer, can improve the performance. These findings highlight the importance of incorporating stochastic elements in the model to achieve better performance. Furthermore, the DGTA model with one set of trained parameters exhibits generalization capability to different instances with varying customer locations and uncertainty levels of travel times without requiring additional training time. Our work contributes to advancing the field of DTSP-TDS, with potential applications in various industrial and logistic problems.

2 - Equitable Routing - Rethinking the Multiple Traveling Salesman Problem

Kaarthik Sundar, Los Alamos National Laboratory, Los Alamos, NM, United States, Abhay Singh Bhadoriya, Deepjyoti Deka

The Multiple Traveling Salesman Problem (MTSP) with a single depot is a generalization of the well-known Traveling Salesman Problem that involves an additional parameter, namely, the number of salesmen. In the MTSP, several salesmen at the depot need to visit a set of interconnected targets, such that each target is visited precisely once by at most one salesman while minimizing the total length of their tours. An equally important variant of the MTSP, the min-max MTSP, aims to distribute the workload among salesmen by requiring the longest tour of all the salesmen to be as short as possible. The min-max MTSP appears in real-life applications to ensure a good balance of workloads for the salesmen. It is known in the literature that the min-max MTSP is notoriously difficult to solve to optimality due to the poor lower bounds its linear relaxations provide. In this paper, we formulate two novel parametric variants of the MTSP called the "fair-MTSP". One variant is formulated as a Mixed-Integer Second Order Cone Program, and the other as a Mixed Integer Linear Program. Both focus on enforcing the workloads for the salesmen to be equitable, i.e., the distribution of tour lengths for the salesmen to be fair while minimizing the total cost of their tours. We present algorithms to solve the two variants of the fair-MTSP to global optimality and computational results on benchmark and real-world test instances that make a case for fair-MTSP as a viable alternative to the min-max MTSP.

3 - Waste Collection Problem with Turn Penalty, Balanced Clustering, and Visual Attractiveness

Seungyeop Lee, Pohang University of Science and Technology (POSTECH), Pohang, Korea, Republic of, Byung-In Kim

The waste collection problem, a variant of the capacitated vehicle routing problem, involves vehicles making multiple waste disposal stops at landfills along their routes. Traditional objectives of waste collection, which focused solely on minimizing transportation costs, are often unsatisfactory to practitioners due to human perception factors. Acceptable solutions prioritize reducing turns, as garbage trucks spend more time maneuvering at intersections. It is also important to balance individual drivers' workloads and minimize route overlap to meet customer preferences of avoiding multiple visits by garbage trucks to the same locations. To address these challenges, we defined measures for a "nice" solution and proposed a cluster-first, route-second algorithm that considers turn penalties, balanced clusters, and visual attractiveness. We evaluate the algorithm's performance using benchmark datasets and real-world waste collection instances to demonstrate its effectiveness and practicality in waste management applications.

4 - Inbound-Outbound Load Plan Adjustments for Terminals in Parcel Delivery Service Networks

Ritesh Ojha, Georgia Institute of Technology, Atlanta, GA, United States, Alan Erera

The load planning problem is a critical challenge in service network design for parcel carriers: it decides how many trailers or loads to operate between terminals in the service network. The base load plan is usually fixed two weeks before the day of operations. Load planners at the terminals then locally modify the base load plan to adjust their outbound loads as the demand forecast changes over time; these local adjustments are often myopic in nature and result in operational inefficiencies. We consider the joint problem for a cluster of terminals where each terminal has planned inbound and outbound trailers and fixed capacity during each sorting periods. Load planners at each terminal need to decide (1) if they want to divert some of the inbound trailers at their terminal to nearby terminals and, (2) to cut some of the planned outbound trailers. The objective is to optimize outbound transportation cost and balance the sortation workload across the terminals in the cluster such that the flow balance, trailer capacity and sorting capacity constraints are satisfied. We propose a mixed-integer program for the inbound-outbound load planning problem. An extensive computational study demonstrates the utility of the model on real-life data for a cluster of three terminals in Atlanta provided by our research partner, a large U.S. parcel carrier. We provide evidence of the efficacy of our solution approach to generate near-optimal solutions that are implementable in practice.

5 - Electric vehicle milk-run scheduling optimization considering time of use pricing

ZHAO ZHAO, Tianjin University, Tianjin, China, People's Republic of, Yida Xu, Zhaofang Mao, Kan Fang

In manufacturing logistics systems, tow trains perform tasks according to a circular delivery, i.e., transporting goods from storage areas to assembly lines and commonly used among manufacturing companies. Unlike traditional vehicle scheduling problems, electric vehicle scheduling considers not only task assignment but also determines the timing and duration of charging. Additionally our study also considering the characteristic of peak-valley-flat electricity pricing in factory electricity usage, discussing electric vehicle scheduling and charging arrangements under time-of-use pricing, as well as balancing workload among vehicles to enhance fairness. To address these issues, an adaptive neighborhood simulated annealing algorithm is developed and compared with commercial solvers and tabu search algorithms. Experimental results demonstrate that the algorithm described in this paper performs well on medium and large instances. Moreover, the study conducts sensitivity analysis on the cost of vehicle usage, revealing insights into reducing the number of electric vehicles employed while maintaining efficiency. By optimizing both task assignment and charging strategies while considering time-of-using pricing and workload balance, manufacturing companies may achieve cost savings and operational improvements in their logistics systems.

MA13

Summit - 333

AI Advancements and Business Systems

Flash Session

Chair: Jiawei Xu, China Agricultural University, Beijing, N/A

1 - Integrating Human-Centric Approaches in Industry 5.0: A Dynamic Capability Perspective

Golnoush Javan, University of Toledo, Toledo, OH, United States, Benjamin George, Steven Wallace

The advent of Industry 5.0 introduces a transformative era in manufacturing, emphasizing a synergistic collaboration between human intelligence, advanced technological capabilities, and intelligent systems. This paper proposes a conceptual framework grounded in Dynamic Capability Theory to examine the adoption of Industry 5.0 from a high-level business perspective. Our study seeks to identify the essential technological and organizational characteristics required for embracing Industry 5.0, with a particular focus on the core principle of human-centricity and its influence on these characteristics. Utilizing systematic literature review and the Gioia methodology for qualitative research, we categorize and analyze various features of Industry 5.0 to construct a comprehensive understanding of the transition requirements. Our findings delineate three major components: technological features, organizational features, and human-centric features, each contributing to the reintegration of humans into the digital manufacturing process. We also explore the implications of these integrations on competitive advantage, emphasizing the role of dynamic capabilities that adapt and evolve in response to rapid technological advancements. This study contributes to the theoretical discourse by providing a structured analysis of how dynamic capabilities can facilitate a smooth transition to Industry 5.0, highlighting the pivotal role of human-centric approaches in enhancing operational effectiveness and innovation in modern manufacturing environments.

2 - AI capabilities and business value

Rajhans Mishra, Indian Institute of Management-Indore, Indore, India

Artificial Intelligence (AI) has become a popular buzzword in business and academia. Generative AI (GenAI) has further triggered a new dimension for the implementation and usage of AI technologies in business. Various techniques like natural learning processing (NLP), deep learning (DL), large language models (LLM), and computer vision are attributed and recognized as AI techniques. Many leading organizations like Microsoft and Google have come up with generic AI platforms (Azure AI & Copilot from Microsoft and Gemini from Google) that can be used by organizations to implement AI capabilities as their solutions. Besides that, several specific AI solutions are also developed and used in specific domains and functional areas. However, there is a lack in terms of the exact capabilities provided by AI solutions and how they can be used to enhance business value. The scope of AI solutions varies based on the capabilities offered, for example, Narrow AI and Strong AI solutions vary significantly based on their scope. This work will explore the various capabilities provided by AI technologies and how they can be mapped to the business value chain in different domains and functional areas. This exploration will be helpful in uncovering the requirements for AI solutions and the development of AI use cases in different business contexts.

3 - Ensuring the Safety of AI Systems

Young Lee, UL Solutions, Melville, NY, United States, Sudhi Sinha

AI systems offer remarkable benefits to society, including enhancements in wellness, convenience, productivity, efficiency, and innovation. However, if misused or malfunctioning, AI can cause significant harm, both unintentional and intentional. These systems are also vulnerable to malicious attacks that can manipulate their functionality or compromise sensitive data. Additionally, improperly trained AI systems can result in unethical outcomes, particularly impacting specific groups of people. Unsafe AI systems pose risks such as physical harm, financial loss, social disruption, and loss of public trust in AI technologies.

This talk introduces a comprehensive framework for evaluating AI application safety. The framework outlines key principles such as reliability, robustness, transparency, explainability, human oversight, fairness, bias mitigation, and data privacy. It defines essential requirements, guidelines, and procedures for each principle, ensuring AI applications align with the highest standards of safety, regulations, and industry best practices. The approach assesses data handling protocols, performance of machine learning algorithms, adaptive training methodologies, decision-making processes, and the overall integrity and reliability of AI systems.

The framework evaluates the conformance of AI systems with international standards like ISO/IEC JTC 1/SC 42 and IEEE-SA, as well as compliance with emerging regulations such as the U.S. AI Executive Order and the EU AI Act. By bridging the gap between rapid technological innovation and safety standard development, this work ensures AI systems are not only innovative and efficient but also safe, ethical, and trustworthy. Establishing AI system safety is vital for building public trust and ensuring the responsible adoption of beneficial AI technologies.

4 - Emotion vs. Information: Understanding the Effect of AI-powered Call Systems on Potential Customers from A Field Experiment

zhe jing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Xin Xu, Yong Jin, Jie Shen

Emerging technologies such as neural networks, cloud computing, and big data have paved the way for the development of artificial intelligence (AI), enabling AI to facilitate business operations. However, AI-human interaction tends to fall short of expectations in real-world settings due to the difference between humans and AI from emotional and informational perspectives. To better understand the mechanism resulting in the different impacts on customer behavior between humans and AI, we conducted a randomized field experiment at a call center within a large securities company. Our study focuses on the differences in customer behavior between AI and human agents when initiating outbound calls to potential customers to encourage them to complete the complex process of opening a financial transaction account. The results show that the AI group and human agent group perform similarly but have different effects on customers from different sources and potential customers who are at different stages of the account opening process due to the differences between emotional support and informational support. The findings of this study contribute to the literature on the application of technology in organizations and provide guidance to organizations on how to implement AI systems effectively.

5 - AIoT Service Adoption and Selling Mode Selection Strategies

Bingxuan Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Jun Lin

With the rapid development of information technology across various fields, many enterprises are seeking to harness Artificial Intelligence of Things (AIoT) technology to enhance their operations and bolster profits. This study constructs a game-theoretic model within the e-commerce domain, involving a manufacturer and an e-commerce platform providing AIoT service. We explore the strategic interplay regarding the adoption of AIoT service by the manufacturer under scenarios where the selling mode is determined either independently by the manufacturer and the platform or both of them.Our focal concerns encompass: (1) whether the manufacturer will adopt the AIoT service provided by the e-commerce platform to transform conventional product into smart one, thereby utilizing consumer data to enhance product utility, and(2) how the information value of the collected data impacts the platform, the manufacturer, and societal welfare. Our study reveals that in certain circumstances, the option for the manufacturer to use the platform's AIoT service, will lead to a better performance of the manufacturer, the platform, and the channel. This insight prompts a reassessment of AIoT integration strategies within the supply chain, offering valuable guidance for e-commerce platforms, manufacturers, and policymakers who aim to optimize societal benefits while maintaining operational excellence.

6 - Untangling the Impact of Artificial Intelligence Innovation on Firm's Market Power

Jiawei Xu, College of Economics and Management, China agricultural university, Beijing, China, People's Republic of, Jianjun Lu, Yi Yang

Artificial Intelligence (AI) technology is reshaping the competitive landscape of the market. Consequently, firms are increasingly focusing on AI-related innovation to enhance their market performance. However, existing research has not fully elucidated the market utility of AI innovation. In light of this gap, this study empirically examines the mechanism by which AI innovation impacts firm's market power, through an analysis of data from Chinese listed companies from 2011 to 2022. The results of regression analysis indicate that AI innovation significantly enhances firm's market power. The impact of AI innovation on market power is moderated by industry heterogeneity, with AI innovation exerting a greater market utility in industries characterized by high concentration and non-technological intensity. The mediation effect test results show that AI innovation can elevate market power by increasing positive media attention, alleviating financial constraints, and reducing internal cost stickiness within firms. Furthermore, the results of response surface analysis reveal that radical AI innovations have a more significant effect on market power than incremental AI innovations. This study not only enriches the theoretical research framework of AI innovation but also provides practical insights and theoretical basis for managers to formulate relevant market strategic decisions.

MA14

Summit - 334

Blockchain Economics

Invited Session

Finance

Chair: Fahad Saleh, University of Florida, Winston-Salem, NC, United States

Co-Chair: Kose John, New York University, New York, NY, United States

1 - How Does Uncertainty Reduction Increase User Welfare in Cryptocurrency Transaction Market? An Empirical Study on Eip-1559

Guangzhi Shang, Florida State Unviersity, Tallahassee, FL, United States, Xi Zhao, Zhichao Wu, Peilin Ai

Transaction fee is fundamental in maintaining the stability and decentralization of a permissionless blockchain, the infrastructure underlying major cryptocurrency networks such as Ethereum and Bitcoin. Like Bitcoin, Ethereum started with a name-your-own-price style fee mechanism. Under this Legacy mechanism, we demonstrate that, surprisingly, only three quarters of the transaction fee is necessary for users to attain the service quality they receive; the remaining one-quarter thus constitutes welfare loss. In August 2021, Ethereum implemented a fee mechanism change. Together, the Legacy methods and the new scheme, which we call Cap-Tip, describe the two most widely used fee mechanisms in all major permissionless blockchains. We explain the innovations embedded in the Cap-Tip mechanism through the lens of information quality and task decomposition. Collecting over 123 million transactions covering the change period, we estimate the causal impact of Cap-Tip on welfare improvement using a doubly robust difference-in-difference estimator that accommodates heterogeneity across timing of and exposure after adoption. The result is an astounding 80% reduction in welfare loss, achieved by users only six days after adoption. Interestingly, we also find that adoption benefits small users more than large ones, alleviating fee disparity and improving fairness. We theorize this difference through user experience and the overconfidence bias. Our study offers a novel operational perspective to unpack how the arguably most significant fee mechanism change in the blockchain space affects user welfare.

2 - Market Fragmentation in Cryptocurrency Exchanges

Wei Chen, University of Connecticut, Stamford, CT, United States, Xiaofeng Liu, Kevin Zhu

This research delves into the repercussions of cryptocurrency exchange market fragmentation. We focus on the effects of new market entrants, both CEXs and DEXs, on the aggressiveness of token pair orders and the resultant price impact. Our findings reveal distinct influences exerted by CEXs and DEXs. While both augment trade volume (in USD) and elevate order submission aggressiveness, the underlying mechanisms diverge. CEXs see an uptick in the number of trades without a corresponding rise in trade size. DEXs, conversely, often report larger individual trades without a surge in their frequency, possibly attributed to substantial blockchain transaction fees and the inherent price predictability due to automated pricing on DEXs. Notably, CEX fragmentation broadens the overall spread, whereas DEXs do not exhibit a significant sway. Furthermore, we find that order aggressiveness increases more for the token pairs with lower pre-entry fragmentation levels, lower trade volume, and linking to non-fiat currencies, while the effects on price impacts are more homogeneous. Moreover, we further extend the analysis to the token level. The results show order aggressiveness is mainly significant at the token pair level, while price impacts from primary tokens also influence associated secondary pairs.

3 - Battle of the Bots: Flash Loans, Miner Extractable Value and Efficient Settlement

Alfred Lehar, University of Calgary, Calgary, AB, Canada, Christine Parlour

Settlement on decentralized ledgers is transparent and batched. The settlement also allows settlement agents to expropriate profitable arbitrage trades. Arbitrage may be socially beneficial or wasteful. We model the effect of an alternate, private settlement on arbitrage. We document payments from arbitrageurs to private settlers that exceed 2 million USD per day.

4 - The Effects of Margin Requirements in Defi Lending Markets

Qiang Wang, Penn State University, State College, PA, United States

This paper investigates the effects of relaxing margin requirements on borrowing decisions in Decentralized Financial lending (DeFi lending) markets. Margin requirements impose leverage constraints by dictating the maximum loan amount relative to the pledged collateral. Using

account-level data, I find that borrowers increase loan balances by 1.6% and collateral holdings by 1.2% in response to a 1% relaxation in margin requirements. Margin requirements disproportionately impact borrowers closer to their leverage constraints. These borrowers take higher leverage following the relaxation. Overall, this paper highlights the role of margin requirements in a credit market without centralized intermediaries.

MA15

Summit - 335

Revenue Management and Marketplace Operations

Invited Session

Revenue Management and Pricing

Chair: Dmitry Mitrofanov, Boston College, Chestnut Hill, MA, United States

1 - Assortment Optimization with Replacement Options for Retail Platforms with Stockout Risk

Yuheng Wang, Cornell University, USA, NY, United States, Dmitry Mitrofanov, Huseyin Topaloglu

There has been a notable surge of online platforms that operate as market facilitators without having direct control over the inventory of the products they sell. This usually leads to potential out-of-stock situations during order fulfillment. Therefore, such online platforms often prompt the customers to pick not only a preferred option, but also a replacement option. We study assortment optimization problems when the customers are offered assortments of products among which they pick their preferred and replacement options. If the preferred option is available, then they are provided with the preferred option. If the preferred option is not available, but the replacement option is, then they are provided with the replacement option. Otherwise, the customers are provided with no products. We use a version of the multinomial logit model to capture the choice process of the customers. We consider two variants of the problem. In the non-adaptive variant, we compute one assortment of products among which the customers pick their both preferred and replacement options. In the adaptive variant, the assortment of products among which the customer chooses her replacement option depends on the preferred option she picked. We show that both variants are NP-hard. Our main technical contributions are a polynomial-time approximation scheme for the adaptive variant and a fully polynomial-time approximation scheme for the non-adaptive variant. Using data from Instacart, we demonstrate that explicitly modeling the presence of a replacement option and using an adaptive strategy to offer assortments for the replacement option can provide significant revenue benefits.

2 - Anti-Competitive Effects of A Dominant Retailer's Guaranteed Profit Margin and Low-Price Contracts

Aditya Jain, Baruch College, Zicklin School of Business, New York, NY, United States, Leela Nageswaran, Haresh Gurnani

A retailer can insist on a guaranteed margin contract with its supplier, wherein they require a margin on each sale. In addition, they may want to have the lowest selling price in a market with competing retailers, and if this strategy of offering the lowest price reduces their margin, the supplier must pay up the difference. There has been recent debate on whether this approach is anti-competitive leading to higher prices and reduced product access. We analyze a game-theoretic model involving a supplier and two retailers where the focal retailer adopts either the guaranteed margin contract, or the guaranteed margin contract with lowest price guarantee. We show that when product category competition is intense, the guaranteed margin contract is anti-competitive: the supplier forms an exclusive distribution with the focal retailer leading to higher prices and reduced product access in comparison with when they both adopt a wholesale price contract. However, if the intensity of competition is low, consumers benefit by paying lower prices with wider access as long as the supplier sets identical wholesale prices, but prices remain higher when they set differentiated wholesale prices. The guaranteed margin contract with lowest price guarantee, however, is always anti-competitive as it leads to higher prices by dampening competition. As such, the ability of the focal retailer to choose the contract type and the intensity of product category competition leads to outcomes that can either benefit or hurt consumers. Our results help inform policy makers on the necessary interventions to alleviate anti-competitiveness.

3 - Personalized Promotion Planning with Strategic Customers

Ozge Yapar, Indiana University, Kelley School of Business, Bloomington, IN, United States, Chloe Glaeser

We study a dynamic personalized promotion problem for a consumer goods retailer facing strategic, repeat customers. We introduce a modified contextual bandit algorithm with a heterogeneous cost term in the reward estimate to reflect how promotions influence future demand. Validated with proprietary data from an online grocer, our approach highlights the need for personalized pricing strategies that consider strategic customer behaviors like stockpiling. Our algorithm, based on Thompson Sampling, offers a computationally efficient method to navigate the complexities of personalized promotion planning. Through numerical simulations using real-world data, we show that considering strategic behaviors through our modified reward function significantly boosts revenue, particularly when many customers are strategic. This work extends the personalized dynamic pricing literature by integrating strategic consumer behavior with effective promotion strategies.

4 - Online Combinatorial Optimization with Group Fairness Constraints

Kumar Kshitij Patel, Toyota Technological Institute at Chicago, Chicago, IL, United States, Negin Golrezaei, Rad Niazadeh, Fransisca Susan

As digital marketplaces and services continue to expand, it is crucial to maintain a safe and fair environment for all users. This requires implementing fairness constraints into the sequential decision-making processes of these platforms to ensure equal treatment. However, this can be challenging as these processes often need to solve NP-complete problems with exponentially large decision spaces at each time step. To overcome this, we propose a general framework incorporating robustness and fairness into NP-complete problems, such as optimizing product rankings and maximizing sub-modular functions. Our framework casts the problem as a max-min game between a primal player aiming to maximize the platform's objective and a dual player in charge of group fairness constraints. We show that one can trace the entire Pareto fairness curve by changing the thresholds on the fairness constraints. We provide theoretical guarantees for our method and empirically evaluate it, demonstrating its effectiveness.

5 - Bundling and Pricing Decisions for Ancillary Products

Tong Xie, The University of Chicago Booth School of Business, Chicago, IL, United States, Rene Caldentey

Service companies are increasingly using ancillary products as a strategy to enhance their core offerings, increase customer satisfaction, and grow their revenue. Broadly speaking, ancillary products are additional products or services offered to consumers alongside their primary purchase. In many cases, there is a separation between the time a customer purchases a core product (e.g., buys an airline ticket or books a hotel room) and the time of its consumption, most ancillary services can be offered and purchased within this time window. This gap introduces an inter-temporal component to Ancillary Revenue Management, simultaneously creating new opportunities and obstacles for firms aiming to implement it effectively.

We study Ancillary Revenue Management in the presence of consumer risk aversion, focusing on scenarios where there is a temporal gap between the purchase and consumption of products or services. We assume that consumers' valuation for an ancillary product is stochastic and evolves over time, becoming more predictable as the time of consumption approaches. We characterize firm's optimal bundling decisions when every ancillary is either sold in a big bundle or sold separately.

Under constant absolute risk aversion (CARA) utility functions, we demonstrate that as consumers become more risk-averse, a firm's optimal policy shifts from selling ancillaries as a bundle to offering them separately. Furthermore, when multiple ancillary products are involved and the valuations of these ancillaries are independent, the optimal bundling policy for each ancillary can be determined independently. We also examine both myopic and strategic purchasing behaviors on the part of consumers.

MA16

Summit - 336

Online Decision Making Algorithms

Invited Session

Revenue Management and Pricing

Chair: Mika Sumida, University of Southern California, Los Angeles, CA, United States

Co-Chair: Jinglong Zhao, Boston University, Boston, MA, United States

1 - A Maximum Likelihood Approach to Randomized Trees

Zhiyuan Sun, UCLA, Los Angeles, CA, United States, Velibor Misic

Randomized trees, unlike traditional deterministic trees, offer a novel framework for predictive modeling by splitting data points with a probability that is determined by their feature vectors and specific parameters at each node. This paper introduces a smooth non-convex estimation problem for learning the parameters of randomized trees from data. We establish that optimizing over randomized trees is theoretically equivalent to optimizing over deterministic trees. We develop generalization guarantees for the log likelihood of randomized trees and proved the convergence of the randomized tree MLE problem. We further showed that the problem is NP-hard. Motivated by this, we develop a heuristic algorithm to efficiently optimize the parameters of randomized trees. Our numerical experiments demonstrate the effectiveness of our method in improving predictive performance over classical top-down induction methods such as CART over a variety of binary and multiclass classification datasets.

2 - Bidding in Uniform Price Auctions for Value Maximizing Buyers

Sourav Sahoo, Massachusetts Institute of Technology, Cambridge, MA, United States, Negin Golrezaei

We study the problem of bidding in uniform price auctions, which are widely used in practice. Although uniform price auctions are non-truthful for bidders with quasilinear utility, empirical findings suggest that this auction format induces truthful bidding. We attribute this difference in theory and practice to the assumption of the behavioral model of the bidders. In this pursuit, we study these auctions in a repeated setting from the perspective of a value-maximizing buyer who aims to maximize their cumulative value across T rounds, subject to per-round return-on-investment (RoI) constraints.

For a RoI-constrained value-maximizing buyer, we study a generalized version of the uniform bidding format, commonly used in practice, which we term as m-uniform bidding. To characterize the optimal m-uniform bid, we introduce and study the notion of universally feasible (UF) bidding policies, which are robust, meaning that RoI feasibility is obtained regardless of the competitors' bids. We show that the optimal class of UF bidding policies is essentially a generalization of truthful bidding policies, which depends only on the valuation curve of the bidder and target RoI. To measure the performance of UF bidding policies against the optimal bidding policy that is not necessarily UF, we introduce a metric called the Price of Universal Feasibility (PoUF) and establish that PoUF<2, irrespective of m and show that the upper bound is tight. We compare the m-uniform bidding interface to the classical uniform bidding (m=1), and show that the total value under m-uniform bidding increases by at most a factor of m.

3 - Online Estimation and Inference for Robust Policy Evaluation in Reinforcement Learning

Yichen Zhang, Purdue University, W Lafayette, IN, United States, Weidong Liu, Jiyuan Tu, Xi Chen

Our work places emphasis on statistical inference for the parameter estimates computed using reinforcement learning algorithms. While most existing analyses assume random rewards to follow standard distributions, we embrace the concept of robust statistics in reinforcement learning by simultaneously addressing issues of outlier contamination and heavy-tailed rewards within a unified framework. In this paper, we develop an online robust policy evaluation procedure, and establish the limiting distribution of our estimator, based on its Bahadur representation. Furthermore, we develop a fully-online procedure to efficiently conduct statistical inference based on the asymptotic distribution. This paper bridges the gap between robust statistics and statistical inference in reinforcement learning, offering a more versatile and reliable approach to policy evaluation.

4 - Diversity-Fair Online Selection

Tongwen Wu, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Yanzhi Li

Selection problems, such as representative democracy, employee recruitment, and task delegation on crowdsourcing platforms, often involve candidates with multiple attributes and unknown arrival. Traditional models, which assume decision-makers (DMs) have complete knowledge of single-attribute value distributions, fall short in these complex real-world scenarios. Our research addresses this gap by focusing on the diversity-fair online selection problem, where the DM must make decisions with limited information and multiple constraints.

We introduce a general online selection model where each candidate's profile is represented by a binary vector across d dimensions, and the selection unfolds over n rounds with a size constraint K. The DM aims to maximize the minimum expected utility across all dimensions, a challenge compounded by the lack of joint distribution knowledge and the need to rely only on marginal information.

Our main contribution lies in developing a linear program-based policy that balances selection probabilities for candidates contributing to multiple dimensions, guided by marginal information. We apply our framework to both a stationary scenario—where an adversary distributes d of ones across candidates in each round—and a general scenario with distinct marginal information. In the stationary case, our method achieves a CR of $\Omega(1/d$^0.75), while in the general scenario, we attain $\Omega(1/(d\log d))$. In addition, we demonstrate that our approach can matches the upper bound for the CR in the general sicario up to a polynomial factor, underscoring the effectiveness of our proposed solutions.

MA17

Summit - 337

Predictive and Constraint Based Learning Techniques with Business Applications

Invited Session

Artificial Intelligence

Chair: Neta Rabin, Tel-Aviv University, Tel-Aviv, Israel

Co-Chair: Gonen Singer, Bar-Ilan University, Ramat-Gan, Israel

1 - Design for Explainability: A New XAI Method Based on Design of Experiments

Irad Ben-Gal , Tel-Aviv University , Tel-Aviv, Israel

We propose a new Design for Explainability (DFX) methodology that for a given trained Machine Learning (ML) model; a feature training set; and a list of potential features interactions generates a Design of Experiments (DOE) plan to measure the importance or the effect of features, while estimating selected high-order features' interactions, with a lower computational cost than the SHAP.Experiments show that DFX and SHAP are equivalent in approximately 80% of the cases in identifying the top 10% of the most affecting features, while the DFX utilize on average 1.05% of the number experiments that are required by SHAP.

2 - The Skill-Fit Model: Utilizing Skills to Advance Machine Learning based Job Recommendation Systems

Hila Chalutz-Ben Gal, Bar-Ilan University, Ramat Gan, Israel

The growing practice of utilizing Machine Learning based Job recommendation systems (JRS) has become a major component in talent management. However, recent disruptions in the labor force (e.g. flexible work, freelancing, gig work) result in the need to adopt a Skill-Fit Model. For this purpose, we present an analysis of the ML based JRS literature and propose a synthesized approach to analyze and model skills. Our results indicate that some JRS features may benefit from the Skill-Fit Model thus may be adopted by ML-enabled JRS in order to achieve improved performance. This study contributes to the understanding and systematically developing skills - based JRS to support the new world of work.

3 - A Hybrid Cost-Sensitive Machine Learning and Optimization Models for the Resource-Constrained Classification Problem

Gonen Singer, Bar-Ilan University, Ramat-Gan, Israel, Danit Shifman, Chen Ben-Mayor, Itay Margolin

Resource-constrained classification tasks are common in real-world applications such as allocating tests for disease diagnosis, hiring decisions when filling a limited number of positions, and defect detection in manufacturing settings under a limited inspection budget. We propose a comprehensive analytic framework for scenarios that, in addition to including multi-class classification problems with misclassification costs, also has constraints on the number of classified samples of classes due to resource limitations. To classify samples under the constraints, the framework uses a probability matrix generated by a trained cost-sensitive classifier as the input for an optimization model with a minimum cost objective and resource allocation constraints. To demonstrate how useful and versatile it can be, the framework utilizing a cost-sensitive neural network and decision tree was implemented in two case studies: one focused on allocating medical resources, and the other on allocating customer service resources. The proposed framework performs significantly better than the alternative common approaches. Our results show that the proposed framework is capable of providing an effective limited-resource allocation for misclassification cost problems.

4 - Spatio-Temporal Time-Series Forecasting Using an Iterative Kernel-Based Regression

Neta Rabin, Tel-Aviv University, Tel-Aviv, Israel

Spatio-temporal time-series forecasting is relevant in various domains like epidemiology or economics. Kernel based regressions, which capture non-linear data-function relationships, provide simple and efficient forecasting models. In this work, we propose a kernel-based iterative regression model that utilizes multiscale kernels to fuse data from various spatial locations and improve the accuracy of time-series forecasting. The model resembles deep learning ideas but is easy to implement and effective for small datasets. Experimental results for forecasting solar energy, epidemiology infections, and fire event forecasts show that the proposed model outperforms existing techniques.

5 - An Approach for classifying Tracheal Stenosis based on EMG signals utilizing an Adaptive Cost-Sensitive Learning method with Asymmetric Misclassification Errors

Sara Naftali, Afeka Tel Aviv Academic College of Engineering, Tel Aviv, Israel, Ohad Volk, Anat Ratnovsky, Gonen Singer

Respiratory disorders, such as asthma, can cause upper airway obstruction, leading to abnormal airway structure. Early detection of this obstruction is crucial in avoiding health deterioration. This study aimed to identify early-stage tracheal stenosis non-invasively by analyzing electromyography (EMG) signals from inspiratory muscles. The classification of tracheal stenosis has been defined as an asymmetric misclassification cost problem. A ResNet-like architecture for tabular data, coupled with an Adaptive Cost-Sensitive Learning (AdaCSL) algorithm, was employed to address this. EMG signals from the external intercostal muscles of four healthy individuals were recorded as they breathed through tubes simulating narrowed and normal airways. Two experimental approaches were developed: one focused on classifying tracheal stenosis in a specific subject by training the model on data from other subjects, mimicking diagnosing a new patient. The other aimed to classify tracheal stenosis by merging data from all subjects to reduce inter-subject variations. The ResNet-like architecture with AdaCSL significantly outperformed other methods, achieving 43%, 48%, and 59% reductions in the first experiment approach across different misclassification cost values. It also demonstrated lower costs in the second approach setting compared to alternative classifiers. These experiments underscore the potential of utilizing EMG signals from inspiratory muscles for respiratory disease diagnosis and highlight the efficacy of the AdaCSL algorithm for personalized monitoring.

MA18

Summit - 338

Analytics for Social Good

Invited Session

Revenue Management and Pricing

Chair: Faidra Monachou, Yale University, New Haven, CT, United States

1 - Implications of Worker Classification in on-Demand Economy

Zhoupeng (Jack) Zhang, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Jianfu Wang

How should gig workers be classified, as employees or as contractors? We study this policy question by highlighting the difference between full- and part-time gig workers and focusing on the welfare of full-timers, who do gigs to earn primary incomes. We develop a multi-period game-theoretic queueing model to study market outcomes in different worker classification scenarios. We show that being classified as independent contractors is not always harmful to full-timers; they can end up with high surpluses thanks to a positive *incentives de-pooling* effect of part-timers' market entry. As such, when all gig workers are reclassified as employees (as per, e.g., Assembly Bill No. 5 in California) and part-timers exit the market, full-timers can be *undercut* (underhired or underpaid) by their platform employer and end up with lower surpluses. When all are reclassified as "contractors^{+}," a UK practice that provides incomplete employee benefits but still gives workers flexibility, additional part-timers incentivized may *overcrowd* the platform and hinder the improvement in full-timers' welfare. We propose two policy alternatives: (1) a Hybrid Mode that classifies full-timers as employees but part-timers as contractors, and (2) a Priority Mode that prioritizes matching full-timers with customers over part-timers. We verify that these differentiated schemes can benefit full-timers as well as other stakeholders in the market. We calibrate the parameters in our model using real-world data and illustrate our insights with simulations of Uber's market over 137 cities in California.

2 - Swap: a New System for Facilitating Resource Sharing Among Nonprofits Through Market Design and Mathematical Optimization

Weixiao Huang, Worcester Polytechnic Institute, Worcester, MA, United States, Carlos Morel Figueroa, Arjun Venat, Elise Deshusses, Yunus Telliel, Sarah Stanlick, Jennifer Pazour, Alexander Teytelboym, Andrew Trapp

Nonprofit organizations (NPOs) face significant challenges in service delivery due to resource scarcity and an inability to scale talents and tangible assets effectively, which is compounded by mismatches between available resources and temporally shifting service demands. Motivated by these observations, we propose a novel resource sharing system, SWAP, what we believe to be the first auction-based resource sharing system designed and implemented in the nonprofit sector. SWAP facilitates resource exchanges through a process of offering-their-own and bidding-upon-others' resources, integrated with virtual currency (re-)distribution mechanisms. The multilateral resource exchange outcomes are determined by integer optimization models accounting for both efficiency and fairness. Empirical and simulated experiments demonstrate that SWAP can address temporal resource needs in the real world. Our custom-built SWAP Hub platform has already facilitated dozens of resource exchanges among tens of pilot NPOs since its initial April 2023 deployment. More NPOs are joining and forming SWAP groups beyond the pilot group, enhancing collective impact on the NPO community and demonstrating strong potential for future expansion.

3 -

Increasing Engagement on Ed-Tech Platforms Through Personalized Nudges: Evidence from the Field

Alex Akira Okuno, New York University, New York, NY, United States, Divya Singhvi, Somya Singhvi

In this study, we collaborate with one of India's leading non-profits focused on catalyzing Early Childhood Education. We run a series of large-scale experiments to evaluate the impact of personalized behavioral nudging using WhatsApp messages. Results demonstrate the significant positive effect that data-driven techniques can have on parents' engagement.

4 - Ecological Inference with a Markov Discrete Choice Model

Charles Thraves, University of Chile, Santiago, Chile

We study a Markov Discrete Choice model in an Ecological Inference setting in a political election context. Specifically, at each ballot box, we observe (i) the total number of votes for presidential candidates and (ii) the total number of votes for house representative candidates' parties, while the number of votes for each combination is unknown. Additionally, each ballot box, depending on the region, has a subset of parties for the House of Representatives. We use the EM algorithm to formulate likelihood maximization with missing data to estimate the model parameters that govern voters' actions. Results from simulated instances show that the original parameters can be accurately identified with the model.

MA19

Summit - 339

Sequential Decision Making Under Uncertainty

Invited Session

Revenue Management and Pricing

Chair: Min-hwan Oh, Seoul National University, Seoul, Korea, Republic of

1 - Nearly Minimax Optimal Regret for Multinomial Logistic Bandit

Joongkyu Lee, Seoul National University(SNU), Gwanak-gu, Korea, Republic of, Min-hwan Oh

In this paper, we study the contextual multinomial logit (MNL) bandit problem in which a learning agent sequentially selects an assortment based on contextual information, and user feedback follows an MNL choice model. There has been a significant discrepancy between lower and upper regret bounds, particularly regarding the maximum assortment size K. Additionally, the variation in reward structures between these bounds complicates the quest for optimality. Under uniform rewards, where all items have the same expected reward, we establish a regret lower bound of $\Omega(d\sqrt{T/K})$ and propose a constant-time algorithm, **OFU-MNL+**, that achieves a matching upper bound of $\tilde{O}(d\sqrt{T/K})$. Under non-uniform rewards, we prove a lower bound of $\Omega(d\sqrt{T})$ and an upper bound of $\tilde{O}(d\sqrt{T})$, also achievable by **OFU-MNL+**. Our empirical studies support these theoretical findings. To the best of our knowledge, this is the first work in the contextual MNL bandit literature to prove minimax optimality --- for either uniform or non-uniform reward setting --- and to propose a computationally efficient algorithm that achieves this optimality up to logarithmic factors.

2 - Improved Algorithms for Multi-period Multi-class Packing Problems with Bandit Feedback

Wonyoung Kim, Columbia University, New York, NY, United States, Garud Iyengar, Assaf Zeevi

We consider the linear contextual multi-class multi-period packing problem (LMMP) where the goal is to pack items such that the total vector of consumption is below a given budget vector and the total value is as large as possible. We consider the setting where the reward and the consumption vector associated with each action is a class-dependent linear function of the context, and the decision-maker receives bandit feedback. LMMP includes linear contextual bandits with knapsacks and online revenue management as special cases. We establish a new estimator which guarantees a faster convergence rate, and consequently, a lower regret in LMMP. We propose a bandit policy that is a closed-form function of said estimated parameters. When the contexts are non-degenerate, the regret of the proposed policy is sublinear in the context dimension, the number of classes, and the time horizon T when the budget grows at least as $T^{1/2}$. We also resolve an open problem posed in Agrawal & Devanur (2016) and extend the result to a multi-class setting. Our numerical experiments clearly demonstrate that the performance of our policy is superior to other benchmarks in the literature.

3 - Lasso Bandit with Compatibility Condition on Optimal Arm

Taehyun Hwang, Seoul National University, Seoul, Korea, Republic of, Harin Lee, Min-hwan Oh

We consider a stochastic sparse linear bandit problem where only a sparse subset of context features affects the expected reward function. To exploit the sparse structure in the unknown reward parameter, existing algorithms for sparse linear contextual bandit use Lasso as an estimator for the reward parameter. The compatibility conditions together with additional diversity conditions on the context features are imposed to achieve regret bounds that only depend logarithmically on the ambient dimension d. In this paper, we demonstrate that even without the additional diversity assumptions, the compatibility condition *only on the optimal arm* is sufficient to derive a regret bound that depends logarithmically on d, and our assumption is strictly weaker than those used in the existing literature. We propose an algorithm that adapts the forced-sampling technique and prove that the proposed algorithm achieves O(*poly*log dT) regret under the margin condition. To our knowledge, the proposed algorithm requires the weakest assumptions among high-dimensional linear bandit algorithms that achieve O(*poly*log dT) regret. Through the numerical experiments, we confirm the superior performance of our proposed algorithm.

4 - Noise-Adaptive Confidence Sets for Linear Bandits and Application to Bayesian Optimization

Kwang-Sung Jun, University of Arizona, Tucson, AZ, United States

Adapting to a priori unknown noise level is a very important but challenging problem in sequential decision-making as efficient exploration typically requires knowledge of the noise level, which is often loosely specified. We report significant progress in addressing this issue in linear bandits in two respects. First, we propose a novel confidence set that is `semi-adaptive' to the unknown sub-Gaussian parameter σ_*^2 in the sense that the (normalized) confidence width scales with $\sqrt{d\sigma_*^2 + \sigma_0^2}$ where d is the dimension and σ_0^2 is the specified sub-Gaussian parameter (known) that can be much larger than σ_*^2. This is a significant improvement over $\sqrt{d\sigma_0^2}$ of the standard confidence set of Abbasi-Yadkori et al. (2011), especially when d is large. We show that this leads to an improved regret bound in linear bandits. Second, for bounded rewards, we propose a novel variance-adaptive confidence set that has a much improved numerical performance upon prior art. We then apply this confidence set to develop, as we claim, the first practical variance-adaptive linear bandit algorithm via an optimistic approach, which is enabled by our novel regret analysis technique. Both of our confidence sets rely critically on `regret equality' from online learning. Our empirical evaluation in Bayesian optimization tasks shows that our algorithms demonstrate better or comparable performance compared to existing methods.

MA20

Summit - 340

Online Learning and Decision-making

Invited Session

Decision Analysis Society

Chair: Sophie Yu, Stanford University, Durham, NC, United States

Co-Chair: Rojin Rezvan, University of Texas at Austin, Austin, TX, United States

1 - Response-Adaptive Clinical Trials with Many Treatments

Jingyi Yan, Duke University, Durham, NC, United States, David Brown

Clinical trials are traditionally designed to be non-adaptive, ensuring the identification of a well-performing treatment by the end of the trial with low bias. However, this approach may result in the inefficient allocation of medical resources to underperforming treatments. In contrast, response-adaptive designs enable the decision maker to alter the number of patients assigned to a treatment given its previous performances. Using an approach based on Lagrangian relaxations, we develop feasible policies and performance bounds. In a simplified model employing beta-binomial belief updates, we establish bounds for the performance loss of the feasible Lagrangian policies and show that these policies are asymptotically optimal when the number of patients and treatments grow at the same rate.

2 - Targeted Policy Learning

Gaoqian Xu, University of Washington, Seattle, WA, United States, Yanqin Fan, Alice Qi

This paper proposes an optimal individualized treatment rule (ITR) targeting the average welfare of a subpopulation of the post treatment outcome population. Operationally we adopt the CVaR of the outcome induced by a binary treatment rule as the objective or welfare function. It interpolates between the average welfare function used in empirical welfare maximization (EWM) and the Rawlsian welfare. It can also be interpreted as distributionally robust EWM allowing for the target population to have a different distribution than the study population. Using the dual formulation of the CVaR welfare function, we propose a debiased estimator of the optimal ITR and establish its regret bounds. In addition, we establish asymptotic normality of the debiased estimator of the optimal welfare facilitating Wald inference. The finite sample performance is studied via simulation and an empirical application is presented.

3 - Reinforced Social Learning

Thanawat Sornwanee, Stanford University, Stanford, CA, United States

I present a knowledge-free mechanism to ensure the zero asymptotic average regret in a sequential social learning framework with binary state of the world and homogeneous preference. This mechanism also ensures that the state of the world can be asymptotically learned, regardless of the choice of Nash equilibrium employed. The asymptotic optimality of the mechanism will also shown to be preserved under different economic scenarios, such as when each agent may be blinded from the history of actions and/or received a utility shock.

4 - A Multi-Dimensional Online Contention Resolution Scheme for Revenue Maximization

Rojin Rezvan, University of Texas at Austin, Austin, TX, United States

We study multi-buyer multi-item sequential item pricing mechanisms for revenue maximization with the goal of approximating a natural fractional relaxation – the ex ante optimal revenue. We assume that buyers' values are subadditive but make no assumptions on the value distributions. While the optimal revenue, and therefore also the ex ante benchmark, is inapproximable by any simple mechanism in this context, previous work has shown that a weaker benchmark that optimizes over so-called "buy-many" mechanisms can be approximable. Approximations are known, in particular, for settings with either a single buyer or many unit-demand buyers. We extend these results to the much broader setting of many subadditive buyers. We show that the ex ante buy-many revenue can be approximated via sequential item pricings to within an O(log2 m) factor, where m is the number of items. We also show that a logarithmic dependence on m is necessary. Our approximation is achieved through the construction of a new multi-dimensional Online Contention Resolution Scheme (OCRS), that provides an online rounding of the optimal ex ante solution. For the welfare objective, constant-factor approximations have been demonstrated for a wide range of combinatorial constraints on item allocations and classes of buyer valuation functions. Our work opens up the possibility of a similar success story for revenue maximization.

MA21

Summit - 341

Advanced Decision Strategies: Integrative Approaches from Machine Learning to Practical Applications

Invited Session

Decision Analysis Society

Chair: Yifan Xie, Adelphi University, Garden City

1 - A Defect Classification Model for Pick and Place Machines using Parallel Neural Network

Yuqiao Cen, Binghamton University, Binghamton, NY, United States

Surface Mount Technology (SMT) involves directly mounting components onto printed circuit boards (PCBs), making it a popular method for low-cost, high-density electronic assemblies. Within the SMT process, the pick-and-place (P&P) step is crucial for placing components after solder paste printing. Automated Optical Inspection (AOI) machines are commonly employed in production lines to detect assembly defects. However, the machines lack the ability to identify the root causes of failures or provide reliable maintenance guidance. With the advent of Industry 4.0, machine learning methods can be applied to improve production line maintenance efficiency. Therefore, the traditional human interact check-out process can be changed into a data-driven and condition-based maintenance process. This paper conducts experiments on initial machine errors and gathers investigative data. It proposes a parallel neural network-based multi-output defect classification model to identify the root causes by leveraging patterns uncovered from experimental data. Compared to traditional root cause identification models, the developed model offers greater flexibility for adjustments.

2 - Enhancing Content Marketing Decisions with Social Media Text Mining: A Case Study of Instagram Fashion Brands

Xiaoru Gao, Northern Illinois University, DeKalb, IL, United States

In order to communicate brand concepts and values to the young generations, many brands are highly active on social media platforms such as Twitter, Facebook, and Instagram. Brand-Generated Content has already become the most common marketing strategy for fashion brands and plays a significant role in influencing consumers' purchase intention. However, not all social media posts successfully yield positive outcomes for brands, as some may attract a notable percentage of negative comments. In this paper, we develop (1) a methodology to detect anomaly of posts that have unusually high percentage of negative user comments and (2) interpret the reason behind anomaly posts. based on Instagram design variables, including semantic text meanings, facial expressions, color scheme and background of photos, and post timing, among others. We collected a data set of brand-generated Instagram posts from ten fashion brands. The data covers the image, text, and user comments posted between 2019 and 2020. Image features were extracted using Convolutional Neural Network, and text topics were generated through Latent Dirichlet Allocation. Our results will help managers design Instagram posts to reduce negative consumer reactions.

3 - Enhancing Feature Interpretability and Reducing Data Costs: Sparse Group PCA with Elastic Regularization

Tianhui Wang, Rutgers University, Piscataway, NJ, United States, Geonseok Lee, Dohyun Kim, Myong Kee Jeong

In modern production environments, the utilization of sensors for monitoring manufacturing processes generates large amounts of data, which offers chances to improve production efficiency and quality by using machine learning models. However, challenges arise in processing raw data and constructing predictive models. Principal component analysis (PCA) is a widely used statistical technique for dimensionality reduction, extracting a low-dimensional subspace in which the variance is maximized (or the reconstruction error is minimized). To improve the interpretability of learned representations, several variants of PCA have recently been developed to estimate the principal components with a small number of input variables. However, existing sparse dimensionality reduction methods only focus on the sparsity in terms of variables or groups of variables, neglecting the simultaneous consideration of sparsity in both variables and groups to enhance feature interpretability and reduce data acquisition costs. To address these challenges, we propose a sparse group PCA method with elastic regularization, integrating sparsity constraints into the objective function to account for both within-group and between-group sparsities. Our approach efficiently uncovers sparse principal loading vectors, eliminating the need for prior assumptions about feature sparsity patterns. We propose an alternating direction method of multipliers for efficient distributed optimization of the non-smooth regularized problem. Experiments conducted on synthetic multi-group data and a semiconductor manufacturing dataset demonstrate the superiority of our proposed method over existing sparse group PCA approaches. Our method exhibits enhanced within-group and between-group sparsities, significantly improving feature interpretability while effectively reducing data acquisition costs.

4 - Optimizing Active Distribution Grids with Weather- and Decision-Dependent Reliability

Gejia Zhang, Rutgers University, Piscataway, NJ, United States, Robert Mieth

The reliability of power systems is critically influenced by weather conditions, which directly affect the operational integrity of physical infrastructure and electricity consumption patterns. Emerging smart grid technology enables system operators to react to temperature fluctuations, but tools to quantify and internalize failure risk on a component-level in operative decisions are still lacking. In this work we develop a distribution system model to optimize the control set-points of distributed generation units, battery energy storage systems, and demand response resources, that explicitly considers weather- and decision-dependent system reliability. Through decision-dependent reliability, our model captures how operational decisions and ambient temperature impact the likelihood of component failures, allowing operators to directly balance cost efficiency and reliability. To address the non-linear and non-convex optimization challenges of the model, we develop an iterative optimization strategy using linearization techniques around initial reliability estimations derived from a baseline scenario. We demonstrate our model and methodology using numerical case studies on IEEE test systems.

5 - Think Like an Organism: Strategies for Low-Surprise Decision-Making via Active Inference

Yulin Li, Rutgers University, New Brunswick, NJ, United States

This presentation explores an innovative strategy for low-surprise decision-making through active inference, inspired by biological organisms. Active inference, based on the free energy principle, posits that organisms minimize discrepancies between internal models and external sensory inputs. This theoretical framework offers a novel approach to developing robust, adaptive algorithms that make meaningful decisions and evaluate the value of those decisions in complex systems, with significant implications for artificial intelligence and automation.

Methodologically, we implement active inference in a simulated robotic control environment. The robot pushes a ball from point A to point B across surfaces of varying friction levels, mimicking dynamically changing conditions. There is no reward unless the ball reaches point B. The robot continuously updates its internal model based on sensory input and makes subsequent decisions to reduce uncertainty. Compared to reinforcement learning approaches, results show that active inference allows the robot to learn and adapt to new conditions more efficiently, requiring fewer trials to train the model, leading to more stable and reliable operations.

The findings underscore the potential of active inference to enhance the adaptability and resilience of autonomous systems across diverse applications, including robotics and autonomous vehicles. By effectively learning world models and making decisions that minimize surprise, active inference offers a promising alternative to conventional RL techniques. Future research will focus on scaling these methods to more complex, real-world scenarios, further validating their practical benefits and applicability.

MA22

Summit - 342

Machine Learning and Causal Inference

Invited Session

Decision Analysis Society

Chair: Sikun Xu, Washington University in St. Louis, St. Louis, United States

Co-Chair: Sikun Xu, Washington University in St. Louis, St. Louis, United States

1 - Impacts of Smart Vending Machines in Different Urban Settings: Tensor Completion with Spatiotemporal Data

Mingrui Zhang, University of Washington- Michael G. Foster School of Business, Seattle, WA, United States, Wanning Chen, Yong Tan, Xusen Cheng

We investigate the treatment effect of deploying smart vending machines (treatment) compared to traditional vending machines (control) across various urban settings. Based on our result, we discuss potential changes in consumer purchasing behavior due to the introduction of technologically enhanced smart vending machines across diverse settings. The dataset we use spans 25 cities, 20 distinct urban settings, and 81 weeks. We employ tensor completion algorithms to predict over 70\% missing sales outcomes in both control and treatment outcomes. Benchmarked against traditional regression models and matrix-based approaches, the tensor completion methods demonstrate superior predictive accuracy. The predictions for the unobserved and non-existent region-setting-time combinations shed light on how to strategically plan and optimize for smart vending machine deployments. More broadly, our analysis underscores the value of leveraging spectral algorithms such as tensor completion to conduct causal analysis on high-dimensional spatiotemporal data, empowering more informed decision-making.

2 - Causal Inference When Controlling for Unstructured Data

Sikun Xu, Washington University in St. Louis, St. Louis, MO, United States, Dennis Zhang, Zhenling Jiang

As unstructured data (e.g., images, and videos) become increasingly available, more empirical research relies on them to derive business and economics insights. Inevitably, this also means that unstructured data have been an important source for empirical researchers to model confounders and control for confounding effects. To control for unstructured data, many studies follow an embedding-then-inference framework: Researchers train a representation learning model to generate lower-dimensional embeddings of the unstructured data, which are then included in the causal inference model. In this paper, we first demonstrate that such a framework can lead to highly biased causal estimates because most representation learning models are not trained to extract relevant confounders from the unstructured data for a particular causal model. We also show that an end-to-end double/debiased machine learning model can avoid arbitrary bias but it requires abundant data to achieve satisfying results. We then propose a debiased embedding framework that effectively and efficiently incorporates unstructured data as confounders. This method extracts low-dimensional embeddings that capture the relevant confounding information from unstructured data and can be easily implemented with other pretrained deep-learning models. Finally, we demonstrate the effectiveness of our proposed method in obtaining accurate causal effect estimates compared to other methods using extensive simulation experiments and a real world dataset.

3 - Learning from User Experience: A Deep Learning Approach to User Embedding for Engagement Prediction and Policy Learning in Online Gaming Platforms

Yu Qin, Arizona State University, Tempe, AZ, United States, Kan Xu, Olivia Liu Sheng

In the competitive online gaming industry, accurately predicting user engagement is essential for maintaining platform vitality and driving strategic decisions. Traditional methods, focusing on demographics and gameplay statistics, fail to fully capture the intricate dynamics of user behavior and interaction. This limitation highlights the need for a more evolved analysis that provides crucial insights into changing engagement and interaction patterns, which are vital for accurate engagement forecasting and effective policy customization. To overcome these challenges, we propose a novel deep learning framework that integrates user match-related data into user embeddings. Our framework leverages diverse data, including game outcomes, in-game dialogue, and paralinguistic elements like emojis, employing multi-task learning to encompass both gaming outcomes and player interactions. This approach offers a comprehensive view of user engagement, addressing the shortcomings of traditional analytics and enhancing the understanding of complex behavior patterns. The anticipated result is an improvement in long-term engagement prediction accuracy and policy learning, enabling targeted promotions and interventions. This research aims to advance user engagement analytics and policy learning, serving as a decision-support tool to optimize business strategies for online gaming platforms.

4 -

An Operational Perspective to Online Content Moderation

Tingrui Shi, University of Chicago, Chciago, IL, United States, Rene Caldentey, Amy Ward, Amir Alwan

The rapid growth of social media users in recent years has led to an increase in the prevalence of harmful content on social media platforms, making online content moderation more important than ever. Content moderation systems nowadays highly rely on artificial intelligence (AI) due to the unprecedented scale of such systems. However, human reviewers are still essential for moderating ambiguous content and providing labels for the training of AI models. We use a Bayesian learning framework to model the AI's classification accuracy as a function of the labeling activities conducted by human reviewers. We derive a diffusion approximation for our model and investigate the operational problem of optimally allocating the capacity of the human reviewers in order to minimize long-term misclassification and holding costs.

MA23

Summit - 343

Multi-objective Optimization: Methods and Applications

Invited Session

Multi Criteria Decision Making

Chair: Ozlem Karsu, Bilkent University, Ankara, Turkey

Co-Chair: Banu Lokman, University of Portsmouth, Portsmouth, United Kingdom

1 - Using Equitable Optimization for the Hazmat Transport Network Design Problem

Özlem Karsu, Bilkent University, Ankara, Turkey, Yunus Çakır

The shipment of hazardous materials is challenging due to the risk that the population centers face during transportation. The policy makers often have concerns for ensuring a balanced allocation of the risks to the population centers. We structure this problem as an equitable optimization problem with multiple objectives aiming to minimize the risk exposure of different neighborhoods. The resulting multiobjective mixed integer linear programming problem is solved to provide the policy makers equitably nondominated options, each with different levels of efficiency (total risk) and fairness (allocation of risk).

2 - Optimising Sampling Plans for MRI-Guided Prostate Biopsies: a Multi-Objective Approach

Banu Lokman, University of Portsmouth, Portsmouth, United Kingdom, Sara Saadatmand, Utku Lokman

We develop a multi-objective optimization-based approach to optimise sampling plans for MRI-guided prostate biopsies. The aim is to generate personalised sampling plans that minimize the number of biopsy cores whilst maximizing the probability of detecting cancer and coverage.

3 - Cutting-Edge Algorithms for Multi-Objective Portfolio Optimization

Pubudu Jayasekara, The University of Tennessee at Chattanooga, Chattanooga, TN, United States, Dilhani Marasinghe

Multi-objective portfolio optimization (MOPO) is an advanced financial strategy that transcends traditional portfolio optimization by integrating multiple investment goals beyond mere return maximization and risk minimization. This approach considers a broader set of factors including liquidity, risk, return, and other relevant investment criteria, reflecting a more holistic view of portfolio management. The development and application of cutting-edge algorithms in MOPO have been propelled by significant enhancements in computational power and algorithmic sophistication. These modern algorithms are adept at managing the intricate trade-offs inherent in diverse investment objectives, making them essential tools in today's complex financial landscape. Consequently, MOPO enables investors to achieve a balanced portfolio that aligns with their multifaceted investment objectives, leveraging state-of-the-art computational techniques to navigate the complexities of modern financial markets.

4 - Aquanutriopt V2.0: A Bi-Objective Multi-Period Optimization Tool for Controlling Harmful Algal Blooms (Habs)

Ashim Khanal, University of South Florida, Tampa, FL, United States, Hadi Charkhgard

We present AquaNutriOpt Version 2.0, an upgraded version of previously published user-friendly Python package designed to tackle a complex combinatorial optimization problem aimed at optimizing nutrient management for the control of harmful algal blooms. This optimization process involves the identification of optimal Best Management Practices (BMPs) and Treatment Technologies (TTs). AquaNutriOpt Version 2.0 boasts two significant enhancements over its predecessor. Firstly, it now accommodates multiple time periods, broadening its scope for comprehensive analysis. Secondly, it empowers users to handle up to two objective functions using a novel custom-built solution approach that factors in diverse budget scenarios and delivers recommendations aligned with the specified objective functions. We demonstrate the efficacy of the package through a real-world case study aiming to control harmful algal blooms in Lake Okeechobee, located in Florida, by simultaneously minimizing both Phosphorous and Nitrogen.

5 - Utilizing Local Upper Bounds for Solving Convex Vector Optimization Problems

Firdevs Ulus, Bilkent University, Ankara, Turkey

There are existing outer approximation algorithms for convex vector optimization problems (CVOPs) that use supporting half Spaces of the upper image to update the polyhedral convex outer approximations. The extreme points of the updated outer approximation are computed by solving a vertex enumeration problem, which gets challenging as the dimension of the objective space increases. On the other hand, local upper bounds (LUBs) are efficiently used to generate approximation algorithms to solve (mixed) integer Mult objective optimization problems. We show that LUB can be used to solve CVOP efficiently under some assumptions on the ordering cone.

MA24

Summit - 344

Optimization Techniques in Network and Scheduling

Contributed Session

Chair: Christopher Montez, OpsLab, Austin, TX, United States

1 - Practical Challenges and Considerations for Network Modeling of a Scheduling Problem

Christopher Montez, OpsLab, Austin, TX, United States

Many scheduling problems may naturally be modeled using a network, with vertices representing a particular event occurring and edges representing a decision of going from one event to another in chronological order. We may then construct an integer program using such a network to find a desirable schedule. However, depending on how the network is constructed, the resulting integer program may prove to be too computationally intense for practical purposes. In this talk, we present our journey for selecting a network for a scheduling problem for the U.S. Air Force. We discuss the various challenges along the way and how we resolved them in order to reduce solve times from the order of hours to seconds.

2 - Stochastic Dynamic Machine Scheduling with Interruptible Set-up Times

Dongnuan Tian, Lancaster University, Lancaster, United Kingdom, Rob Shone, Kevin Glazebrook

We consider a problem in which a machine is scheduled dynamically within a network to process jobs at different demand points. Costs are accumulated as jobs wait to be processed. The time needed for a machine to travel between two nodes in the network represents the “set-up time” needed to switch from processing one type of job to another. By using a network formulation, we can model complex relationships in switching times between different types of activity and also allow the switching times to be interrupted. The problem can be formulated as a Markov decision process in which arrival times, service times and switching times are uncertain and the objective is to minimize the expected long-run average cost. However exact solutions using dynamic programming are not possible due to the complexity of the state space. Heuristics for certain special cases of the problem have been proposed in the literature and in this talk we discuss how to adapt these heuristics to our problem. We also discuss how forward-thinking strategies can be used to develop improved heuristics and show the results obtained by applying these methods to networks of various different configurations.

3 - An outer approximation method for finding exact solutions to the discrete network design problem

Michael Levin, University of Minnesota, Minneapolis, MN, United States, David Rey

The problem of network design under user equilibrium route choice behavior appears in many specific contexts, but is difficult to solve because user equilibrium is itself a convex program resulting in a bi-level optimization problem. We specifically focus on discrete network design focusing on binary decisions to open or not open new links. Despite its introduction by Leblanc in 1975 and broad relevance, progress on exact solution methods has been limited. We develop a path-based outer approximation method for the system optimal discrete network design problem, and extend it to the bi-level problem with user equilibrium route choice. A path-based approach is used because link-based formulations do not scale well with network size. The idea behind outer approximation is to solve a sequence of relaxed master problems with system optimal traffic assignment that approximate the convex objective function by linear underestimators from points corresponding to subproblem solutions obtained at integer link design solutions. The subproblems are system optimal and user equilibrium traffic assignment problems with fixed design variables. Outer approximation is applicable because when the integer decision variables are fixed, network design becomes a convex traffic assignment problem. The system optimal problem is used to obtain lower bounds, whereas upper bounds are obtained through user equilibrium traffic assignment. Branching on relaxed binary design variables is used, and system optimal-based lower bounds are used to eliminate some branch nodes.

4 - Ant Colony Optimization for Path Planning in Search and Rescue Operations

Michael Morin, Department of Operations and Decision Systems, Université Laval, Québec, QC, Canada, Irène Abi-Zeid, Claude-Guy Quimper

In search and rescue operations, an efficient search path is a path maximizing the probability of finding survivors. Maximizing this objective quickly enough and with sufficient realism can have a substantial positive impact in terms of human lives saved. In this presentation, we address the problem of efficiently optimizing search paths in the context of the optimal search path problem with visibility, based on search theory. To that end, we develop and evaluate ant colony optimization algorithm variants where the goal is to maximize the probability of finding a moving search object with Markovian motion, given a finite time horizon and finite resources (scans) to allocate to visible regions. Our empirical results, based on evaluating 96 variants of the metaheuristic with standard components tailored to the problem and using realistic size search environments, provide valuable insights regarding the best algorithm configurations. Furthermore, our best variants compare favorably, especially on the larger and more realistic instances, with a greedy heuristic and a state-of-the-art mixed-integer linear program solver. With this research, we add to the empirical body of evidence on the configuration and application of ant colony optimization algorithms, and pave the way for search path optimization in operational decision support systems for search and rescue.

5 - Proving an Approximation Ratio by Mathematical Programming

Insoo Park, POSTECH, Pohang, Korea, Republic of, Kyungduk Moon, Kangbok Lee

The identical parallel machine scheduling problem, aimed at minimizing the makespan, is a central topic in scheduling theory. Among the various approximation algorithms, MULTIFIT stands out for its significance and theoretical impact. Despite being proposed approximately 45 years ago, the approximation ratio of the MULTIFIT algorithm remains unproven for specific numbers of machines. In this work, we introduce a new methodology that utilizes mathematical programming-based analysis to determine the approximation ratio for these unresolved cases. Our approach not only addresses this long-standing question but also offers a promising framework for analyzing other algorithms that can be encoded as mathematical programming models.

MA25

Summit - 345

Gurobi/Nextmv

Invited Session

Technology Showcase

1 - One Goal, Multiple Paths: Understanding Gurobi's Python APIs for Efficient Model Construction

Maliheh Aramon, Gurobi Optimization, North Vancouver, BC, Canada

Discover the versatility of Gurobi’s Python APIs in our workshop, “One Goal, Multiple Paths.” We’ll explore three key approaches to model building:

1. Term-based modeling involves constructing models with individual variables and constraints for detailed control.
2. Matrix-based expressions utilize linear algebra for efficient, large-scale model construction.

3. Data-first methods leverage pandas DataFrames and Series, integrating seamlessly with data-centric workflows.

In this workshop, we will walk you through the principles, best practices, and practical examples for each method. By the end of the workshop, you will be equipped to choose and implement the most suitable approach for your optimization projects. Join us to master Gurobi's powerful Python APIs and enhance your modeling skills.

2 - The What, why, and how of DecisionOps: It's not just about what you run, but how you run it

Carolyn Mooney, Nextmv.io Inc, Philadelphia, PA, United States

Optimization is founded upon the promise of efficiency and improving solutions. While it's common to believe success starts and ends with an algorithm's runtime, there are tremendous gains to be had beyond those bounds.

While continued optimization model and solver development is important, the next chapter of optimization will be defined by how teams deploy, test, and operate these solutions. Whether you implement a trustworthy sedan or F1 racer of optimization tech, there are key (and often overlooked) efficiencies to be realized in how teams test, tune, troubleshoot, and operate their solutions.

Operations research and decision science teams balance the demands of budgets, headcount, revenue, timelines and success metrics, building and improving DecisionOps practices will save time, money, and headache in pursuit of optimization efficiency. This session will explore how through real-world applications and storytelling.

MA26

Summit - 346

Advanced Network and Optimization Techniques for Resilient Systems

Contributed Session

Chair: Sarper Aydin, Harvard University, 150 Western Ave, Allston, MA, 02134, United States

1 - Building Back Better: Decentralized Recovery Modeling in Sociotechnical Systems using Strategic Network Dynamics

Negin Maddah, Northeastern University, BOSTON, MA, United States, Babak Heydari

The ability of sociotechnical systems to recover from disruptions through decentralized, bottom-up processes is a crucial element of their overall resilience, which is also among the most overlooked in resilience research. This study proposes a computational framework leveraging dynamic multi-agent networks to conceptualize, model, and quantify decentralized recovery processes building upon strategic network formation methods. This innovative approach addresses limitations in conventional strategic network formation models by accounting for the fluidity of structural shifts induced by the strategic behaviors of individual agents, and by acknowledging the inherent heterogeneity within agent populations. The central objective of this research is to highlight the critical role of bottom-up recovery and present a systematic, adaptable framework for modeling these processes in diverse real-world scenarios. We achieve this through four key elements: modeling post-disruption responses, exploring their interdependencies, accommodating multiple equilibria, and generating metrics for scenario comparison and forecasting. Our dynamic strategic network formation model offers valuable insights into crucial trade-offs in decentralized recovery processes and facilitates the integration of both bottom-up and top-down resilience strategies. Furthermore, it could serve as a predictive tool informing initial network architectural decisions by estimating likely post-disruption equilibrium states based on the initial network structure. Ultimately, this study provides a robust platform for analyzing network efficiency and resilience trade-offs, setting the stage for more informed decision-making in complex sociotechnical systems.

2 - Federated Optimization of the Integrated Transportation Distance by Dual Subgradient Methods

Zhengqi Lin, Rutgers University, Piscataway, NJ, United States, Andrzej Ruszczynski

The Integrated Transportation Distance is a novel metric in the space of stochastic kernels. We identify its usefulness for addressing several federated learning problems, such as the Wasserstein barycenter problem under the privacy-preserving setting and federated clustering. We propose efficient federated dual subgradient algorithms for optimizing the Integrated Transportation Distance without violating the privacy restrictions. Owing to the absence of any matrix-vector operations, the algorithms exhibit very low complexity of each iteration and significant scalability. We illustrate their virtues on several examples.

3 - Learn-to-Construct Cuts for Nested Benders Decomposition in Multi-stage Stochastic Network Optimization Problems

Shima Mohebbi, George Mason University, Fairfax, VA, United States, Babak Aslani, Ran Ji

The temporal propagation of uncertainty and the sequential decision-making nature of large-scale network optimization problems in different application areas, such as flow-based restoration of interdependent infrastructure networks, have not been investigated extensively from a multi-stage stochastic formulation point of view in the literature. While Nested Benders Decomposition (NBD) has emerged as a promising solution algorithm to tackle such multi-stage stochastic problems, the method suffers certain computational bottlenecks and fails to be considered a scalable solution algorithm. Having shed light on these computational deficiencies, we utilize machine learning algorithms to embed accelerating components into the NBD framework to provide a scalable solution algorithm for large-scale network optimization problems. Specifically, we collect the training data from the NBD and train a fast regression with a customized loss function to construct outer-approximated Bender's cuts. In addition, we use an end-to-end learning component for the forward pass of NBD to reduce the computational time. Even though we use approximated Benders cuts, the proposed framework maintains the exactness of the NBD, as we solve the sub-problems and the master problem in the backward pass with such cuts in an exact approach. We evaluate the proposed Learn-to-

Construct Cuts framework in randomly generated networks and a case study to show the computational improvements in terms of the number of cuts, the number of iterations, and the convergence rate.

4 - Reduced-Basis Method for Fast Solutions of Traffic Flow Problems

Ali Pirmohammadi, New York University, New York, NY, United States, Daniel Vignon

Obtaining repeated and/or real-time solutions of traffic flow equations in a network is computationally prohibitive, especially when historical information on traffic flow and travel demand patterns is not suitable (e.g: disaster response scenarios). Even though model order reduction methods exist to address this issue, their application is limited in this context due to the hyperbolic nature of the partial differential equations (PDEs) describing the system. In this study, we investigate the reduced-basis (RB) method for fast solution of first-order macroscopic traffic flow models both at the link and network level. The RB approximates the solution manifold of the PDE system with a linear basis, allowing rapid construction of the solution for given time horizon, boundary, and initial conditions. This is accomplished through an offline and an online stage. In the offline stage, we obtain the reduced basis functions through an effective greedy sampling method. We overcome the inherent reduction challenge for hyperbolic PDEs by relying on cumulative density, rather than density, for the construction of the basis. In the online stage, we solve a specific instance of the problem with the aid of the Empirical Interpolation Method (EIM). Applied at the link and network levels, our method shows promising computational improvement.

5 - Multi-Agent Resilient Consensus under Intermittent Faulty and Malicious Transmissions

SARPER AYDIN, Arizona State University, Tempe, AZ, United States, Orhan Eren Akgun, Stephanie Gil, Angelia Nedich

In this work, we consider the consensus problem in which legitimate agents share their values over an undirected communication network in the presence of malicious or faulty agents. Different from the previous works, we characterize the conditions that generalize to several scenarios such as intermittent faulty or malicious transmissions, based on trust observations. As the standard trust aggregation approach based on a constant threshold fails to distinguish intermittent malicious/faulty activity, we propose a new detection algorithm utilizing time-varying thresholds and the random trust values available to legitimate agents. Under these conditions, legitimate agents almost surely determine their trusted neighborhood correctly with geometrically decaying misclassification probabilities. We further prove that the consensus process converges almost surely even in the presence of malicious agents. We also derive the probabilistic bounds on the deviation from the nominal consensus value that would have been achieved with no malicious agents in the system. Numerical results verify the convergence among agents and exemplify the deviation under different scenarios.

MA27

Summit - 347

Models of Entrepreneurship and Innovation

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Tian Chan, Emory University, Atlanta, GA, United States

Co-Chair: Zhi Chen, National University of Singapore, Singapore, Singapore

1 - A Market for Lemons? Guidelines for A Vigilant Application of Artificial Intelligence in Entrepreneurship Research

Moren Levesque, York University, Toronto, ON, Canada, Martin Obschonka

The rapid expansion of AI applications is transforming entrepreneurship scholarship. Due to potential information asymmetries, especially between authors and reviewers, the field could suffer from unproductive knowledge exchanges. Notably, if newly published, low-quality articles supported by AI go undetected in academic journals, they could create a ‘market for lemons.’ If the field can avoid such unintended and inaccurate knowledge exchange, then the AI revolution could potentially usher in a new golden era for the field. We discuss actions necessary to elevate the field to a higher level of AI resilience, while steadfastly updating its foundational principles and core values.

2 - Meet Our Friendly Virtual Assistant! the (Many) Hurdles to Chatbot Adoption in Customer Service

Brett Hathaway, Brigham Young University, Provo, UT, United States, Evgeny Kagan, Maqbool Dada

Despite recent advances in large language models, chatbot technology continues to face adoption hurdles. We report the results of four studies (a retrospective survey and three incentivized experiments) that examine the choice between the chatbot channel and the live agent channel in customer service. We show that people respond positively to improvements in chatbot performance; however, the chatbot channel is utilized less frequently than expected time minimization would predict. This underutilization is caused by two separate behaviors: gatekeeper aversion (aversion to any service format that may involve a transfer to a different server) and algorithm aversion (aversion to an algorithmic service provider). Among the two factors, gatekeeper aversion is the more dominant one, accounting for 73% to 86% of behavior. In contrast, algorithm aversion is a weaker factor, and is only present for those aged 40 and older in our data. We further unpack gatekeeper aversion and show that it is driven in roughly equal parts by two factors: risk aversion and acclimation (an aversion to obstacles that slow down progress towards service completion).

3 - How Funding Incentives at Lean Startups Distort Product Experiments and Innovation

Onesun Yoo, UCL School Of Management, University College London, London, United Kingdom, K Sudhir, Zihou Zhou

The lean startup method (LSM) is an experimentation-driven approach to learn about product-market fit for improving startup success, but it abstracts away from the financing incentives of the entrepreneur. This paper models the LSM process in the presence of an investor that seeks to learn about entrepreneur's private type while the entrepreneur seeks to learn about the consumer. We show that the investor cannot infer the entrepreneur's type based on the entrepreneur's actions, and therefore must rely on the product test outcome to do so. This can create an incentive for the high-type entrepreneur to distort the experimentation strategy. Our analysis identifies two possible types of distortions relative to the LSM without financing need: one that prioritizes funding at the expense of learning, and the other that prioritizes minimizing the risk of false positive test outcome at the expense of leaving potential innovation ideas wasted. Overall, the information asymmetry about the entrepreneur's type leads to misallocation of funds, wherein the low-type entrepreneurs may get funded and under low investor confidence when the high-type entrepreneurs may not be funded. A noteworthy implication from our model is that entrepreneurs facing negative investor stereotypes (e.g., women, minority entrepreneurs) endogenously sacrifice learning for funding, launching overall less successful ventures and obtaining lower payoffs than those not facing such negative stereotypes. Thus, investor stereotypes on entrepreneurs can become self-reinforcing, accentuate payoff inequality and hurt welfare.

4 - Learning by Failing: the (Unintended) Consequence of Test Reporting on Autonomous Vehicle Training

Zhi Chen, National University of Singapore, Singapore, Singapore, Wenjie Xue

The success of autonomous vehicles (AVs) depends on its access to various real-world driving scenarios. To improve AVs' algorithms, firms send testing vehicles to discover unknown scenarios. While these scenarios are valuable for AVs' future improvements, they trigger failures. We study how a test reporting regulation affects firms' testing strategies.

5 - Experience and Pivoting in Entrepreneurial Product Development

Janne Kettunen, The George Washington University, Washington, DC, United States, Karthik Ramachandran

Pivoting is a significant shift from the current product under development to an alternative opportunity. Pivoting is strongly advocated in the popular press and commonly used in entrepreneurial product development. However, not all pivots are fruitful. Several practical examples and our conversations with early-stage entrepreneurs suggest that a high fraction of pivots do not succeed. We explore the entrepreneurs' experience level, which manifests itself in reduced overconfidence, as a mechanism that drives the pivoting decision. We aim to answer how the entrepreneurs' experience level impacts their pivoting decisions. We develop a decision-theoretic model to analyze the entrepreneurial choice regarding whether to pivot or persist while accounting for the overconfidence and market and pivot opportunity uncertainties. We show that lack of experience, i.e., higher level of overconfidence, can result in systematic overutilization or underutilization of pivoting. Furthermore, we show that novice (more overconfident) entrepreneurs are likely to make radical risky pivots, whereas experienced (less overconfident) entrepreneurs are more likely to make safe pivots. Our results suggest how, depending on the experience level of the founders and their likely pivoting behavior, entrepreneurs should set the duration of their supply chain contracts, choose the type of employees to hire, and select mentors. For government and educators, our results indicate what type of training programs to provide for novice and experienced entrepreneurs. For venture capitalists, the results imply different approaches for evaluating startups with novice and experienced founders.

MA28

Summit - 348

Leveraging Research Centers for Sustainable Operations

Panel Session

MSOM: Sustainable Operations

Co-Chair: Karen Donohue, University of Minnesota, Minneapolis, MN, United States

1 - Moderator Panelist

Karen Donohue, University of Minnesota, Supply Chain & Operations, Minneapolis, MN, United States

2 - Panelist

Suvrat Dhanorkar, Pennsylvania State University, University Park, PA, United States

3 - Panelist

Beril Toktay, Georgia Institute of Technology, Atlanta, GA, United States

4 - Panelist

Owen Wu, Indiana University, Bloomington, IN, United States

MA30

Summit - 421

Quantum Computing and Optimization (Part 1)

Invited Session

OPT: Linear and Conic Optimization

Chair: Zeguan Wu, Lehigh University, Bethlehem

Co-Chair: Mohammadhossein Mohammadisiahroudi, Lehigh University, Bethlehem, PA, 18015, United States

1 - A quantum dual logarithmic barrier method for linear optimization

Zeguan Wu, Lehigh University, Bethlehem, PA, United States

Quantum computing has the potential to speed up some optimization methods. One can use quantum computers to solve a linear system of equations via the Quantum Linear System Algorithm (QLSA). Quantum Linear System Algorithms (QLSAs) can be used as an oracle for algorithms that require solving a linear systems of equations at each iteration. The dual logarithmic barrier method is used to solve linear optimization problems. This method , at each iteration, requires the solution of a linear system to compute Newton directions. In this paper, we use QLSA to solve the linear system at each iteration of the dual logarithmic barrier method. Due to the noise in contemporary quantum computers, the output of QLSA is noisy. Hence, we need to work with the inexact variant of the dual logarithmic barrier method. Thus, this paper proposes an inexact feasible dual quantum logarithmic barrier method for linear optimization. Convergence properties of the method are studied and we show that this method has $\mathcal{O}(\sqrt{n}L)$ iteration complexity, where L is the bit length of the input data.

2 - Arbitrary state preparation via quantum walks

Igor Gaidai, The University of Tennessee Knoxville, Knoxville, TN, United States, Alvin Gonzales, Rebekah Herrman, Colin Campbell, Ji Liu, Teague Tomesh, Zain Saleem

In this talk, I will introduce the concept of continuous time quantum walks on dynamic graphs and show how these walks can be used to prepare an arbitrary quantum state in O(nm), where n is the number of qubits and m is the number of non-zero amplitudes in the state. Then I will show how the walks can be converted to a circuit model and demonstrate performance comparison between different methods of state preparation.

3 - Quantum Algorithms and Software for Nonconvex Continuous Optimization

Xiaodi Wu, University of Maryland, College Park, College Park, MD, United States, Jiaqi Leng, Ethan Hickman, Joseph Li, Yufan Zheng, Samuel Kushnir, LEI Fan

Identifying efficient quantum algorithms for nonconvex optimization would be appealing for both theoretical and practical applications. A conventional approach to achieving quantum speedups in optimization relies on the quantum acceleration of intermediate steps in classical algorithms while maintaining the overall algorithmic trajectory and solution quality unchanged. We propose Quantum Hamiltonian Descent (QHD), which is derived from the path integral of dynamical systems corresponding to certain classical optimization algorithms, as a genuine quantum approach to nonconvex optimization. Specifically, we prove that QHD can efficiently solve a family of nonconvex continuous optimization instances, each characterized by exponentially many local minima. Meanwhile, a comprehensive empirical study suggests that representative state-of-the-art classical optimization algorithms/solvers, including Gurobi, would require super-polynomial time to solve such instances. We also propose several implementation techniques that enable us to run QHD on today's analog quantum computers for solving large-scale non-convex quadratic programming (QP) problems. Based on the QHD algorithm and its real-machine implementation, we have developed a user-friendly software package named QHDOPT, specifically designed for nonconvex optimization.

4 - The quantum central path algorithm: improvements and extensions

Brandon Augustino, JPMorgan Chase, New York, NY, United States

We propose a novel quantum algorithm for solving linear optimization problems by quantum-mechanical simulation of the central path. While interior point methods follow the central path with an iterative algorithm that works with successive linearizations of the perturbed KKT conditions, we perform a single simulation working directly with the nonlinear complementarity equations. Combining our approach with iterative refinement techniques, we obtain an algorithm that avoids dependence on condition number and exhibits polylogarithmic dependence on precision. We then discuss how our approach can be extended to semidefinite optimization and convex quadratic optimization problems.

5 - Quantum optimization with local counterdiabatic driving

Jiayu Shen, JPMorgan Chase, New York City, NY, United States, Changhao Li, Ruslan Shaydulin, Zichang He, Marco Pistoia

Recent advancements in quantum computation including quantum control have significantly broadened its applicability in addressing complex optimization problems. One of the key challenges in quantum optimization is the preparation of a target ground state with high fidelity from an easy-to-prepare ground state. Traditionally, quantum adiabatic evolution, the standard technique used for such tasks, requires slow driving, which translates to impractically long evolution times for many applications. This work introduces the application of local counterdiabatic (LCD) driving as an efficient shortcut to adiabaticity, aimed at overcoming these limitations in quantum ground state preparation. Focusing on applications including portfolio optimization and the transverse-field Ising model, our results reveal that the proposed LCD driving protocols can substantially outperforms traditional adiabatic techniques with simple accessible quantum control. This demonstration of LCD methods paves the way for preparing the ground state of complex systems using readily available controls. Our study highlights the potential of local control-based quantum shortcut to adiabaticity to advance quantum optimization by providing faster and more efficient solutions to real-world problems.

MA31

Summit - 422

Interplay Between Optimization, Statistics, and Learning - Part I

Invited Session

OPT: Machine Learning

Chair: Lijun Ding, University of California San Diego, La Jolla, CA, 77843, United States

Co-Chair: Liwei Jiang, Georgia Institute of Technology, Atlanta, GA, United States

1 - Reinforcement Learning in Latent Environments: Online Guarantees via off-Policy Evaluation

Jeongyeol Kwon, Wisconsin Institute for Discovery, Madison, WI, United States, Yonathan Efroni, Mannor Shie, Constantine Caramanis

In many real-world decision problems there is partially observed, hidden or latent information that remains fixed throughout an interaction. Such decision problems can be modeled as Latent Markov Decision Processes (LMDPs), where a latent variable is selected at the beginning of an interaction and is not disclosed to the agent initially.
In last decade, there has been significant progress in designing learning algorithms for solving LMDPs under different structural assumptions. However, for general LMDPs, there is no known learning algorithm that provably matches the existing lower bound~\cite{kwon2021rl}. We effectively resolve this open question, introducing the first sample-efficient algorithm for LMDPs without *any additional structural assumptions*.
Our result builds off a new perspective on the role off-policy evaluation guarantees and coverage coefficient in LMDPs, a perspective, which

has been overlooked in the context of exploration in partially observed environments. Specifically, we establish a novel off-policy evaluation lemma and introduce a new coverage coefficient for LMDPs. Then, we show how these can be used to derive near-optimal guarantees of an optimistic exploration algorithm.
These results can be of interest beyond the LMDP class, and especially, for partially observed environments.

2 - 1-bit Matrix Completion for Causal Panel Data Models under Binary Outcomes

Jianzhi Xu, University of Southern Carlifornia, Los Angeles, CA, United States

In this paper, our main focus is to introduce 1-bit matrix completion to address the problem of predicting average and individual causal effects when dealing with binary outcome panel data. The core idea of this approach is to adopt a convex optimization approach to recover the control/treated potential outcome matrix based on the observed 1-bit (binary) outcomes. By doing so, we can more accurately predict causal effects under such a binary setting.

3 - Proximal Random Reshuffling Under Local Lipschitz Continuity

Xiaopeng Li, Columbia University, New York, NY, United States, Cedric Josz, Lexiao Lai

We study proximal random reshuffling for minimizing the sum of locally Lipschitz functions and a proper lower semicontinuous convex function without assuming coercivity or the existence of limit points. The algorithmic guarantees pertaining to near approximate stationarity rely on a new tracking lemma linking the iterates to trajectories of conservative fields. One of the novelties in the analysis consists in handling conservative fields with unbounded values.

4 - Algorithms for Budget-constrained D-optimal Design

Weijun Xie, Georgia Institute of Technology, Atlanta, GA, United States, Jiaqi Wang, Ilya Ryzhov

D-optimal experimental design is a classical statistical problem in which one chooses a collection of data vectors, from some available large pool, in order to maximize a measure of predictive quality. In the classical formulation, the only constraint is on the cardinality of the collection, i.e., the number of vectors chosen. We study a more general budget-constrained variant in which vectors have heterogeneous costs, and develop four new algorithms (two deterministic, and two randomized) with approximation guarantees. Our methods handle heterogeneous costs using a novel exchange rule that interchanges packs of data vectors whose total costs are similar (up to some controlled amount of rounding error). The algorithms outperform the only existing method for this problem from both theoretical and empirical standpoints.

MA32

Summit - 423

Recent Advances in Online Decision Making

Invited Session

OPT: Optimization Under Uncertainty

Chair: Jiashuo Jiang, Hong Kong University of Science and Technology, Hong Kong, N/A

Co-Chair: Yilun Chen, CUHK Shenzhen, Shenzhen, N/A

1 - High-Dimensional Prediction for Sequential Decision Making

Georgy Noarov, University of Pennsylvania, Philadelphia, PA, United States, Ramya Ramalingam, Aaron Roth, Stephan Xie

We study the problem of making predictions of an adversarially chosen high-dimensional state that are unbiased subject to an arbitrary collection of conditioning events, with the goal of tailoring these events to downstream decision makers. We present efficient algorithms for solving this problem, as well as a number of applications that stem from choosing an appropriate set of conditioning events.

For example, we can efficiently make predictions targeted at polynomially many decision makers, giving each of them optimal swap regret if they best-respond to our predictions. We generalize this to online combinatorial optimization, where the decision makers have a very large action space, to give the first algorithms offering polynomially many decision makers no regret on polynomially many subsequences that may depend on their actions and the context.

Further, we develop a novel transparent alternative to conformal prediction for building valid online adversarial multiclass prediction sets. We produce class scores that downstream algorithms can use for producing valid-coverage prediction sets, as if these scores were the true conditional class probabilities. We show this implies strong conditional validity guarantees including set-size-conditional and multigroup-fair coverage for polynomially many downstream prediction sets. Moreover, our class scores can be guaranteed to have improved L2 loss, cross-entropy loss, and generally any Bregman loss, compared to any collection of benchmark models, yielding a high-dimensional real-valued version of omniprediction.

2 - Online Local False Discovery Rate Control: a Resource Allocation Approach

Hongyu Chen, Massachusetts Institute of Technology, Cambridge, MA, United States, Ruicheng Ao, David Simchi-Levi, Feng Zhu

We consider the problem of sequentially conducting multiple experiments where each experiment corresponds to a hypothesis testing task. At each time point, the experimenter must make an irrevocable decision of whether to reject the null hypothesis before the next experimental result arrives. The goal is to maximize the number of discoveries while maintaining a low error rate at all time points measured by local False Discovery Rate (FDR). We formulate the problem as an online knapsack problem with exogenous random budget replenishment. We start with general arrival distributions and show that a simple policy achieves a $O(\sqrt{T})$ regret. We complement the result by showing that such regret rate is in general not improvable. We then shift our focus to discrete arrival distributions. We find that many existing re-solving heuristics in the online resource allocation literature, albeit achieve bounded loss in canonical settings, may incur a $\Omega(\sqrt{T})$ or even a $\Omega(T)$ regret. With the observation that canonical policies tend to be too optimistic and over claim discoveries, we propose a novel policy that incorporates budget safety buffers. It turns out that a little more safety can greatly enhance efficiency --- small additional logarithmic buffers suffice to reduce the regret from $\Omega(\sqrt{T})$ or even $\Omega(T)$ to $O(\ln^2 T)$. We extend the policy to the scenario with continuous arrival distributions as well as time-dependent information structures. We conduct both synthetic experiments and empirical applications on a time series data from New York City taxi passengers to validate the performance of our proposed policies.

3 - Stochastic subgradient descent for high-dimensional stochastic control

umur cetin, Cornell, ithaca, NY, United States, Yilun Chen, David Goldberg

We employ a convex optimization framework and demonstrate that stochastic subgradient descent leads to a simulation-based PTAS for optimal stopping and multiple stopping, as well as certain generalizations thereof. Our approach establishes novel connections between high-dimensional stochastic control, convex optimization, and simulation, enabling PTAS for a broad range of high-dimensional control problems. Our work builds on Chen and Goldberg's recent iterative simulation-based PTAS for optimal stopping.

4 - Optimistic Q-learning For Average Reward and Episodic Reinforcement Learning

Priyank Agrawal, Columbia University, New York, NY, United States, Shipra Agrawal

We present an optimistic Q-learning algorithm for regret minimization in average reward reinforcement learning under an additional assumption on the underlying MDP that for all policies, the expected time to visit some frequent state s_0 is finite and upper bounded by H. Our setting strictly generalizes the episodic setting and is significantly less restrictive than the assumption of bounded hitting time {\it for all states} made by most previous literature on model-free algorithms in average reward settings. We demonstrate a regret bound of $\tilde{O}(H^5 S\sqrt{AT})$, where S and A are the numbers of states and actions, and T is the horizon. A key technical novelty of our work is to introduce an $\overline{L}$ operator defined as $\overline{L} v = \frac{1}{H} \sum_{h=1}^H L^h v$ where L denotes the Bellman operator. We show that under the given assumption, the $\overline{L}$ operator has a strict contraction (in span) even in the average reward setting. Our algorithm design then uses ideas from episodic Q-learning to estimate and apply this operator iteratively. Therefore, we provide a unified view of regret minimization in episodic and non-episodic settings that may be of independent interest.

MA33

Summit - 424

Efficient Methods for Large-scale Optimization Using Inexactness I

Invited Session

OPT: Optimization Under Uncertainty

Chair: Robert Baraldi, Sandia National Laboratories, Seattle, WA, 87123, United States

Co-Chair: Johannes Milz, Georgia Institute of Technology, Atlanta, GA, United States

1 - Structure Utilization for Reducing Dimensional Dependency for Nonsmooth Derivative-Free Optimization

Ching-pei Lee, Institute of Statistical Mathematics, Tokyo, Japan, Geovani Grapiglia

Deterministic methods for derivative-free optimization that use finite difference to approximately conduct first-order updates usually have an evaluation cost per iteration proportional to the problem dimension, making these methods prohibitively expensive for high-dimensional problem.
In this work, we consider a regularized setting of minimizing the sum of a smooth function and a nonsmooth regularization term, where we only have access to the function value of the smooth term but assume full knowledge of the regularization. By identifying the structure at the point of convergence induced by the regularization using an inexact proximal gradient method constructed through finite difference, we propose an acceleration method whose per-iteration evaluation cost locally depends on the rank of the structure instead of the original problem dimension, and thus making the algorithm much more efficient without harming the convergence guarantees. We discuss how the structure can be found under a general error bound condition, and cover related convergence properties.
We further consider a progressive identification strategy that gradually decreases the estimated rank of the structure through an approximate Riemannian proximal gradient method by confining to the manifold that represents the structure at the current point, and discuss how it could potentially improve the convergence speed.

2 - A Stochastic Trust Region Method for Nonsmooth Optimization

Aurya Javeed, Sandia National Laboratories, Albuquerque, NM, United States

This talk is about minimizing a smooth term plus a convex nonsmooth term. We present a stochastic proximal Newton trust region algorithm that assumes models and estimates of the objective are sufficiently accurate, sufficiently often. Like previous stochastic trust-region algorithms such as STORM, which handles stochastic optimization with random models, we use martingale theory to prove our algorithm is globally convergent with probability one.

3 - Risk-averse optimization using randomized quasi-Monte Carlo methods

Olena Melnikov, Georgia Institute of Technology, Atlanta, GA, United States, Johannes Milz

We establish epigraphical and uniform laws of large numbers for sample-based approximation. Our sample-based approximation schemes include Monte Carlo (MC), quasi-Monte Carlo (QMC), and certain randomized quasi-Monte Carlo integration (RQMC), such as scrambled net integration. Our results can be applied to the approximation of risk-averse stochastic programs and stochastic variational inequalities. Finally, we empirically compare the performance of MC, QMC, and RQMC approximations for high-dimensional risk-averse stochastic optimization problems. We find that RQMC approaches based on Halton and Sobol sequences yield smaller variance than MC and QMC methods.

4 - Adaptive Randomized Sketching for Nonsmooth Dynamic Optimization

Robert Baraldi, Sandia National Laboratories, Seattle, WA, United States

Dynamic optimization problems are used for control and modeling of many applications, including optimal flow control, full waveform inversion, and medical imaging. Notably, these applications are plagued by significant computational challenges; namely, memory is a limiting factor on problem size. Evaluation and storage of the entire state trajectory is required for both cost function and derivative computation. Additionally, nonsmooth regularizers or auxiliary constraints on the optimization variables that arise via physical or user-desired features can prohibit usage of traditional smooth optimization techniques. We introduce a trust-region algorithm for minimizing the sum of a smooth, nonconvex function and a nonsmooth, convex function that addresses these two challenges. Our trust-region algorithm employs randomized sketching to store a compressed version of the state trajectory and adjoints, and adaptively learns the rank of the state sketch. We prove convergence of our method with near optimal memory requirements and demonstrate the efficacy of our method on PDE-constrained optimization problems.

MA34

Summit - 425

Advances in Continuous Optimization Algorithms

Invited Session

OPT: Nonlinear Optimization

Chair: Tianjiao Li, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Zhe Zhang, Georgia Tech, Atlanta, GA, United States

1 - Smoothing approach for nonconvex-nonconcave minimax problems

Yangyang Xu, Rensselaer Polytechnic Institute, Troy, NY, United States, Hari Dahal, Wei Liu

We consider nonconvex-nonconcave minimax problems and present a smoothing framework to solve the problem. Convergence results and numerical results will both be presented. Compared to existing works, our algorithm can work without assuming concavity-like structure or PL-like condition on the dual part.

2 - Hyperparameter Tuning via Trajectory Predictions: Stochastic Prox-Linear Methods in Matrix Sensing

MENGQI LOU, Georgia Institute of Technology, ATLANTA, GA, United States

Iterative algorithms are the workhorses of modern signal processing and statistical learning, and are widely used to fit large-scale, complex models to random data. While the choice of the hyperparameters of an algorithm determines both the efficiency and fidelity of the learning pipeline, it is common for this choice to be made heuristically, usually either guided by upper bounds on error convergence or by expensive trial-and-error.

Motivated by these issues, we study an important iterative stochastic algorithm——the mini-batched prox-linear method——for a canonical nonconvex model-fitting problem—rank-1 matrix sensing with random data. We derive a low-dimensional deterministic recursion that predicts the error of this method and show, using a non-asymptotic framework, that this prediction is accurate for any batch-size and a large range of step-sizes. In particular, our analysis reveals that this method, though stochastic, converges linearly from a local initialization with a fixed step-size to a statistical error floor. Our analysis also exposes how the batch-size, step-size, and noise level affect the convergence rate and the eventual statistical estimation error. Notably, our convergence guarantees are sharp, in the sense that we provide matching upper and lower bounds on the error trajectory. We also demonstrate how to use our deterministic predictions to perform hyperparameter tuning (e.g. step-size and batch-size selection) without ever running the method.

3 - Policy Dual Averaging for General State and Action Space

Caleb Ju, Georgia Institute of Technology, Atlanta, GA, United States, Guanghui Lan

We first present a substantial generalization of the recently developed policy mirror descent method to deal with general state and action spaces. We introduce new approaches to incorporate function approximation into this method, so that we do not need to use explicit policy parameterization at all. Moreover, we present a novel policy dual averaging method for which possibly simpler function approximation techniques can be applied. We establish linear convergence rate to global optimality or sublinear convergence to stationarity for these methods applied to solve different classes of RL problems under exact policy evaluation. We then define proper notions of the approximation errors for policy evaluation and investigate their impact on the convergence of these methods applied to general-state RL problems with either finite-action or continuous-action spaces. Finally, our presentation includes preliminary numerical results of policy dual averaging applied to various problems in optimal control and sequential decision making.

4 - Conditional gradient methods for smooth functional constrained optimization

Yuyuan Ouyang, Clemson University, Clemson, SC, United States, Yunheng Jiang, Zhe Zhang

We propose new projection-free algorithms for solving functional constrained optimization problems. For convex functional constrained optimization problems, we show that a straightforward extension of conditional gradient (CG) method, called the constrained CG (CCG) method, can solve an ε-approximate solution while the gradient evaluation complexity and the linear objective optimization complexity are both of order $O(\varepsilon^{-1})$. Further, by incorporating a sliding procedure in CCG, we show it could obtain a better $O(\varepsilon^{-1/2})$ gradient evaluation complexity while maintaining the $O(\varepsilon^{-1})$ linear objective optimization complexity. Moreover, we show that CCG with a linesearch strategy can also be adapted for solving nonconvex functional constrained optimization problems. If the objective function is nonconvex but the constraints are convex, our proposed method has an $O(\varepsilon^{-2})$ complexity to compute an approximate stationary point. If both the objective and constraint functions are nonconvex, our proposed method has an $O(\varepsilon^{-2})$ complexity to compute an approximate Fritz-John point.

MA35

Summit - 427

Amazon Flexible Under-the-Roof Capacity Planning

Invited Session

Transportation Science and Logistics (TSL)

Chair: Roman Levkin, Amazon, Seattle, WA, United States

1 - Associate Capacity Prediction

Esra Sisikoglu, AMAZON, Seattle, WA, United States, Jeronimo Callejas

Effective short-term labor and capacity planning at Amazon hinges on accurately predicting the capacity to flex-up or flex-down the labor supply. From hiring decisions at the planning level to shift offering at the execution level, all models leverage predicted associate capacity to optimally match demand with labor supply. This presentation discusses the design of an capacity prediction service that identifies the associate capacity – how much we can flex up or down – within a given time horizon, at a given surge rate, with a given associate pool. The strength of this design lies in its flexibility to provide predictions for various use cases at different granularity, enabling coordination between planning and execution levers.

2 - Labor Capacity Management with Surge

Dev Das, Amazon.com, Seattle, WA, United States

Amazon uses additional surge pay incentive to associates to accept shifts where the demand of associate hours is high and acceptance rate of shifts is low. Optimizing surge pay recommendations involves understanding associate responsiveness to surge pay that requires measuring the acceptance rate of surge shifts for various surge amounts, the time when shifts start, and how close to the shift start time is the surge offered. This presentation will discuss scientific approaches developed to address this problem. The approaches range from cost-neutral, fair-pay-adherent randomized control experimental design to fully-automated continuous learning approaches using multi-armed bandit models that are used across multiple Last Mile operations.

3 - Integrating Associate Preference into Labor Planning: Sentiment Analysis

Rachel Rutkowski, Amazon, Seattle, WA, United States, Jeronimo Callejas, Roman Levkin, Thomas Fillebeen, Martin Savelsbergh

Labor planning often relies on capacity predictions generated from observable labor data. These data are useful in predicting broad capacity trends but are insufficient in capturing the nature, directionality, and magnitude of the causal capacity relationship between the labor context and associate labor behavior. Thus, we leverage associate listening tools (e.g., surveys) to capture associate-generated sentiment data that are synthesized to recover latent factors that inform associate labor behavior. These data are used to refine model specification selection (e.g., complexity, assumptions, features) which results in more durable models that capture the broader, nuanced labor context within which associates operate.

4 - Integrating Associate Preference into Labor Planning: Structural Modeling

Jeronimo Callejas, Amazon, Seattle, WA, United States, Rachel Rutkowski, Roman Levkin, Thomas Fillebeen, Martin Savelsbergh

Labor planning often relies on capacity predictions generated from observable labor data. These data are useful in predicting broad capacity trends but are insufficient in capturing the nature, directionality, and magnitude of the causal capacity relationship between the labor context and associate labor behavior. Thus, we leverage associate listening tools (e.g., surveys) to capture associate-generated sentiment data that are synthesized to recover latent factors that inform associate labor behavior. These data are used to refine model specification selection (e.g., complexity, assumptions, features) which results in more durable models that capture the broader, nuanced labor context within which associates operate.

5 - Enhancing Associate Capacity via Shift Reservations

Evan Wyse, Amazon.com, Inc., Seattle, WA, United States

While Amazon's flexible workforce claims the majority of their shifts proactively, the company has recently begun to experiment with pushing specific opportunities out to associates. A propensity-matching experiment indicates that associates who are given specific, guaranteed shift opportunities work more hours on average, as well as working more hours at night. This presentation will discuss specific algorithms used to identify the shifts that associates would be most likely to claim using a recommender system. Further, we will also articulate the methods used to ensure that shift opportunities are allocated fairly and optimally to all associates.

6 - Amazon Flexible Under-the-Roof Capacity Planning

Thomas Fillebeen, Amazon, Seattle, WA, United States, Roman Levkin

Overview of flexible Under-the-Roof (UTR) capacity planning across Amazon operations, its unique challenges and opportunities for Science to make impact. The objective is to minimize hiring, attrition, productivity loss and costs while maintaining capacity reliability and introducing capacity flexibility to structurally respond to changes in customer demand. To accommodate demand volatility cost effectively, many high-speed fulfillment operations started with slack capacity. However, as this specialty business matured and capacity constraints became top of mind, these businesses started shifting towards maximizing throughput. In this environment, it is crucial to predict capacity to adapt and respond to disruptive changes and deploy labor execution levers. This in turn requires a holistic approach - from machine learning to qualitative research.

MA36

Summit - 428

Dynamic Decision-Making in Last-Mile Fulfillment

Invited Session

TSL: Freight Transportation

Chair: Sara Reed, University of Kansas, Lawrence, KS, United States

Co-Chair: Dipayan Banerjee, Loyola University Chicago, Chicago

1 - Continuous-TIME Dynamic Yield Management: near-Optimal Policies and Applications to Multichannel Fulfillment

Dipayan Banerjee, Georgia Institute of Technology, Atlanta, GA, United States, Alan Erera, Alejandro Toriello

Motivated by multichannel retailers using in-store inventory to fulfill both in-store demand and rapid delivery requests from online customers, we study the finite-horizon continuous-time dynamic yield management problem with stationary arrival rates and two customer classes. We consider a class of linear threshold policies proposed by Hodge (2008), in which each online (i.e., less desirable) customer is accepted if and only if the remaining inventory at the customer's arrival time exceeds a threshold that linearly decreases over the selling horizon. Using a discrete-time Markov chain representation of sample paths over inventory-time space, we show that a range of such linear threshold policies achieve uniformly bounded regret. We then generalize this result to analogous policies for the same problem with arbitrarily many customer classes. Numerical simulations demonstrate linear threshold policies' competitiveness with existing heuristics and illustrate the effects of the linear threshold's slope. We discuss other potential applications of these results to rapid e-retail fulfillment.

2 - Smart Parking in Last-Mile Delivery

Sara Reed, University of Kansas, Lawrence, KS, United States, Michael Lash

Parking poses a challenge in last-mile delivery, particularly in urban environments. One way to mitigate this challenge may be to provide the driver with real-time parking information to assist with driver decision-making. We introduce the Dynamic Parking Delivery Problem with Real-Time Information (DPDP-I) to model en-route decisions regarding where the delivery person should park and which customers to serve from each parking spot. We formalize the DPDP-I as a Markov Decision Process and develop a dynamic solution approach. Our computational results shows that our heuristic policy successfully leverages historical and real-time parking information as well as the interdependencies between parking availability and customer density to reduce route completion times.

3 - In-Store Order Picking for E-Grocery

Charlotte Köhler, European University Viadrina, Frankfurt (Oder), Germany, Jan Fabian Ehmke, Ann Campbell

E-grocers struggle to create profitable business models, and careful planning is needed to balance tight margins with high customer expectations. Although numerous studies emphasize optimizing customer acceptance and efficient delivery, the costs and resources required for picking orders are less examined despite their significant influence on profitability.

In this presentation, we delve into in-store picking costs within the e-grocery context and introduce a cost evaluation function designed to give retailers a precise assessment of the resources needed for picking orders, thereby assessing feasibility and maximizing the number of accepted orders. For e-groceries, the time taken to pick items is a critical determinant of costs and use of resources, as longer picking times directly translate to increased labor expenses and fewer orders that can be picked.

We are conducting a comprehensive assessment of the proposed cost evaluation function, utilizing real-world data to ensure accuracy and relevance for different order picking strategies. Our assessment employs a detailed analysis of store layouts like those of REWE in Germany, combined with historical data from a former German e-grocer involving over 400,000 order baskets. This data provides insights into typical order compositions and the resources needed for various basket types, enhancing our understanding of critical costs in e-grocery picking. The insights from this study will provide valuable input for a comprehensive analysis of both picking and delivery processes in the e-grocery sector.

4 - Gamifying the Vehicle Routing Problem with Stochastic Requests

Justin Goodson, Saint Louis University, Saint Louis, MO, United States, Nicholas Kullman, Nikita Dudorov, Martin Cousineau, Jorge Mendoza

Do you remember your first video game console? We remember ours. Decades ago, they provided hours of entertainment. Now, they’ve been repurposed to solve dynamic and stochastic optimization problems. With deep reinforcement learning techniques posting superhuman performance on a wide range of Atari games, we consider the problem of representing a classic logistics problem as a game. Then, we train agents to play it. We consider several game formats for the vehicle routing problem with stochastic requests. We show how various design features impact agents’ performance, including perspective, field of view, and minimaps. With the right game design, general purpose Atari agents outperform optimization-based heuristics, especially as problem size grows.

MA37

Summit - 429

Emerging Topics in Transportation and Logistics

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Hai Wang, Singapore Management University, Singapore, Singapore

1 - Pricing Shared Rides

Yifan Shen, University of Washington, Seattle, WA, United States, Chiwei Yan, Julia Yan

Propelled by high rates of personal car ownership (91.5% of American households in 2020, Davis and Boundy 2022) and low vehicle occupancies (1.54 passengers per car in 2017), shared rides have been envisioned as a means of transporting compatible riders together in a carpool-type service, thereby reducing vehicle miles and carbon emissions. However, major ridesharing platforms have long struggled to maintain a healthy and profitable shared rides product. To understand why shared rides have struggled, we analyze procedures commonly used in practice to set static prices for shared rides, and discuss their pitfalls. We then propose a pricing policy that is adaptive to matching outcomes, dubbed match-based pricing, which varies prices depending on whether a rider is dispatched alone or to what extent she is matched with another rider. Analysis on a single origin-destination setting reveals that match-based pricing is both profit-maximizing and altruistic, simultaneously improving cost efficiency (i.e., the fraction of cost saved by shared rides relative to individual rides) and reducing rider payments relative to the optimal static pricing policy. These theoretical results are validated on a large-scale simulation with hundreds of origin-destinations from Chicago ridesharing data. The improved efficiency of match-based pricing drives a reduction in vehicle miles while also benefiting riders and platforms. In addition, these effects are especially pronounced under high-cost, low-demand settings, expanding access to shared rides where they have historically been most challenging.

2 - The Value of Flexibility in Last-Mile Delivery

Haoqing Wang, Department of Logistics & Maritime Studies, Faculty of Business, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Sheng Liu, Shuaian Wang

In on-demand last-mile delivery, it is common to assign each driver to serve a designated area, which enhances the driver's familiarity with the region but limits the overall flexibility of the delivery system. By analyzing the real-world delivery data provided by a food delivery platform, we identify imbalanced capacity among regions. Motivated by this phenomenon, we investigate the value that driver flexibility can bring to last-mile delivery. With the objective of improving service level, i.e., reducing the expected wait time of customers, we consider three mechanisms: (i) full flexibility between two (neighboring) regions, (ii) one-side partial flexibility between two regions, and (iii) two-side partial flexibility between two regions. We find that (a) full flexibility can improve the service level to the greatest extent; (b) partial flexibility also enhances the service level, and under optimal configurations, it can achieve a significant proportion of the service level improvement attained by full flexibility. Moreover, we analyze the impact of region size on the results and expand the analysis to multiple orders and regions.

3 - Navigating Algorithmic Disparity in Ride-Hailing: Examining Spatial-Temporal Disparate Impacts of Transportation Network Company Algorithms in New York City

Hanyong Xu, Massachusetts Institute of Technology, Cambridge, MA, United States

With the rise of Transportation Network Companies (TNCs) like Uber and Lyft, examining their pricing and matching algorithms becomes crucial. Previous studies have shown that these algorithms can discriminate against certain demographic groups. However, the specific spatial-temporal distributions and the associated demographic characteristics have not been thoroughly explored. Understanding the development of such algorithmic discrimination is essential. It enables platform users to be aware of potential pricing volatility and helps decision-makers advocate for better industry regulations.

Under these conditions, the presentation will cover three parts. First, it demonstrates that there are spatial disparities in pricing and service availability caused by the algorithms of Transportation Network Companies (TNCs) in New York, while controlling for socio-demographic factors and trip characteristics. Second, it reveals how these disparities and their relationship to socio-demographic factors have evolved over the course of the pandemic. Lastly, the presentation introduces a method that leverages socio-demographic awareness in the demand prediction model to improve prediction fairness.

4 - Assessing and Explaining Urban Transportation Network Efficiency

Hai Wang, Singapore Management University, Singapore, Singapore

Urban transportation networks are critical infrastructures ensuring the mobility of people and goods while having a significant influence on the environment. To assess the service quality and environmental impact of current and emerging transportation services, a critical challenge for service providers and policymakers is to measure and understand the efficiency of urban transportation networks. Existing transportation network efficiency metrics either do not address the efficiency of service trips beyond two demand points or do not provide large-scale efficiency assessments on real networks. In this paper, leveraging network optimization insight and modern transportation data, we introduce a network efficiency metric capable of assessing the efficiency of multi-demand vehicle routing based services on urban transportation networks at scale, while accommodating different service modalities and demand distributions. Furthermore, to understand what makes a network efficient, we identify primary micro/macroscopic network properties explaining the network efficiency. By deploying our methodology, we construct a new dataset to measure and explain network efficiency for 200 cities worldwide. Through case studies, we show that our method can assess the network efficiency under diverse transportation service modalities. Specifically, we examine the efficiency of multimodal urban transportation services, the time evolution of the efficiency of New York City's subway network, and the efficiency of a last-mile delivery service.

5 - Multi-Criteria Decision Making in Transportation Planning

Hongyi Jiang, Cornell University, Ithaca, NY, United States, Samitha Samaranayake

In transportation planning, decision-making often involves navigating complex tradeoffs between various competing objectives. In this talk, I will discuss the development and application of multi-criteria decision-making methods to address some of the key challenges in this domain. The discussion will cover the theoretical foundations and practical implementations of optimization models and algorithms designed to balance factors such as efficiency, cost, sustainability, and reliability in transportation systems. The insights gained from our work can inform ongoing and future research in advancing multi-criteria decision-making approaches, both in transportation and other areas facing similar complexities.

MA38

Summit - 430

Air Transport Operations

Invited Session

Aviation Applications

Chair: Jan-Rasmus Kunnen, Wing Labs, Hamburg, Germany

1 - Deriving strategic ATM regulations from network-optimal routing via decision trees

Jan-Rasmus Kunnen, Wing Labs, Hamburg, Germany

We use decision tree methodology to develop applicable and effective ATFM regulations for strategic planning of air traffic flows in Europe in times of capacity shortages.

The methodology was developed as part of an industrial project to upgrade the planning tools for the European Network Manager. It consists of 3 parts: First, we determine the best flight-to-route assignment for all flights given capacity restrictions and pre-determined trajectory options. Second, use decision trees to derive individual regulations (RAD measures) from the resulting trajectories by aggregating individual trajectories on selected criteria (Origin, Destination, sector entry time). Finally, we combine the individual measures for each Area Control Center (ACC) by selecting the most effective set of non-overlapping measures.

We apply the method on a real-life case study of 10,000 flights, where we determine the most effective ATFM measures for a week-long capacity reduction in two Paris ACCs in November 2024.

2 - TBA

Ruhollah Heydari, Delta Air Lines, Inc., Atlanta, GA, United States

3 - The Airport of Future: Prescriptive AI Integration for Optimizing Delta Air Lines' Worldport Operations

Ash Omidvar, Delta Air Lines, Inc., Atlanta, GA, United States, Ruhollah Heydari, Dayana Cope

Efficient airport operations hinge upon seamless coordination among diverse resources, both above and below the wing, to ensure on-time and high-quality flight experience. The envisioned "Airport of Future" endeavors to amalgamate subject matter expertise with advances in artificial intelligence (AI) technologies to attain optimal operational outcomes. In this presentation, the Delta Decision Science team elucidates the deployment of prescriptive AI methodologies at the Worldport to enhance the efficiency of both above-wing (crew and customer-related) and below-wing (ground operations-related). Through a synergistic integration of advanced algorithm and real-time data analytics, the practiced AI framework facilitates proactive decision-making, resource allocation, and scheduling optimizations. By harnessing the power of prescriptive AI, the Airport of Future further elevates operational performance, enhances customer experience, and ensures safety and welfare for team Delta.

4 - Distributionally Robust Integer Programming and Air Traffic Management

Max Li, University of Michigan, Ann Arbor, MI, United States, Haochen Wu, Alexander Estes

The airport ground holding problem seeks to minimize flight delay costs due to reductions in the capacity of airports. However, the critical input of future airport capacities is often difficult to predict, presenting a challenging yet realistic setting. Even when capacity predictions provide a distribution of possible capacity scenarios, such distributions may themselves be uncertain (e.g., distribution shifts). To address the problem of designing airport ground holding policies under distributional uncertainty, we formulate and solve the airport ground holding problem using distributionally robust optimization (DRO). We address the uncertainty in the airport capacity distribution by defining ambiguity sets based on the Wasserstein distance metric. We propose reformulations which integrate the ambiguity sets into the airport ground holding problem structure, and discuss dicretization properties of the proposed model. We discuss comparisons (via numerical experiments) between ground holding policies and optimized costs derived through the deterministic, stochastic, and distributionally robust airport ground holding problems. Our experiments show that the DRO model outperforms the stochastic models when there is a significant difference between the empirical airport capacity distribution and the realized airport capacity distribution. We note that DRO can be a valuable tool for decision-makers seeking to design airport ground holding policies, particularly when the available data regarding future airport capacities are highly uncertain.

5 - Natural Disaster-Resilient Spaceport Network Planning

Haochen Wu, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

In recent years, the United States commercial space launch sector has experienced unprecedented growth, driven by companies such as SpaceX and Blue Origin. Geohazards, such as landslides, flooding, and erosion have long caused challenges to transportation systems in the US. In particular, hurricanes have caused significant challenges to spaceport operations at Cape Canaveral Space Force Station and Kennedy Space Center. Our motivation aligns with the strategic goals of the USSF to address the problem of creating a spaceport network that not only handles the demand of space access but is also resilient for natural disasters that will inevitably arise. Given historical data of natural disasters' annual frequencies (such as those obtained from, we apply the Chance-Constrained SPFL model (CC-SPFL). The idea behind the CC-SPFL is that we generate samples of the random variable based on the annualized frequency data, and for at least some percentage of all samples the demand of all types of missions are satisfied. We will also apply importance sampling technique into CC-SPFLP to deal with rare but significant events such as earthquake, and thus generating potential spaceport locations under uncertainties of different types of natural disasters.

6 - Flight Timetabling under Context-dependent Choice Model

Yufeng Cao, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

Flight timetabling impacts an airline's operating profit. We consider a joint flight timetabling and revenue management problem with passengers choosing among multiple itineraries. The classic multinomial logit (MNL) choice model does not well depict the interaction between parallel itineraries. So, we introduce a context-dependent MNL model where the attraction value of one product is contingent on the offer set. The context-dependent MNL model extends common choice models and provides alternative explanations to several empirical observations. We formulate the joint flight timetabling and revenue management problem as a mixed integer programming model. We present a two-stage decomposition framework to compute the network timetable within a practical time budget. Numerical results suggest that the combination of the formulation and the solution approach leads to profit improvement for airlines.

MA39

Summit - 431

Operations Research and Analytics at BNSF Railway

Invited Session

Railway Applications

Chair: Andrea Arias, BNSF Railway, Keller, TX, United States

1 - Real time optimization and sequential decision making in railroad hump yard operations

Avnish Malde, BNSF Railway, Fort Worth, TX, United States, Paul Kuhn

The hump yards are the large merchandise terminals in the railroad that assist in classifying cars belonging to common outbound blocks. Every hump yard consists of sub-yards, i.e., arrival yard, classification bowl, and departure yards. The arriving railcars that must be classified are stored on arrival yard tracks. The outbound block for every railcar that needs classification is known. Every railcar requiring classification is then sent over the hump, also known as the hill, according to the hump order of the arrival yard tracks and the sequence of the railcar on the arrival yard track. Classification tracks are preassigned for every outbound block that needs to be built. The switches and retarders aid in routing the railcar to the correct classification track. When all the required cars of the outbound trains are humped and its build time has approached, the train is built by gathering all the railcars parked on corresponding classification tracks. The cleared classification tracks can now be assigned to construct a new outbound block. This research studies the multi-period sequential decision-making problem for optimal outbound block assignment to classification tracks by considering the operational constraints of every sub-yard so that we maximize the overall performance of the hump terminal. We model this problem as a multi-objective program and solve it using a combination of hierarchical and blended objective approaches.

2 - Multi-period network optimization for railroad intermodal networks

Hedieh Ashrafi, Southern Methodist University, Dallas, TX, United States, Paul Kuhn

This project introduces an integrated recommendation tool for *Condition-Based Gate Management* (CBGM) throughout BNSF intermodal network. The proposed recommendation system is comprised of two steps. First, we develop a mathematical formulation to optimally orchestrate the intermodal unit flow such that the total number of penalized blocked ingates is minimized subject to yard inventory, lift (ramp and deramp) and offspot capacities. The proposed model connects major operations including ingate, ramp, pipeline, deramp and outgate within and between the intermodal hubs. A solution to the model involves the number of units to ingate, allocate, ramp, not-ramp, deramp, offspot and outgate at origin, destination, domestic/internation and load/empty levels over an 11-day planning horizon. In the second stage, we design a what-if algorithm to automatically drill down into the obtained solution and identify the underlying reasons for possible undesirable events such as allocations and offspots.

3 - Dual-stream, time-space network optimization for resource allocation at railroad intermodal hubs

Vishal Badyal, BNSF Railway, Fort Worth, TX, United States, Dasaradh Mallampati

Railroad intermodal hub facilities are logistical marketplaces where customers drop off individual units (i.e. in-gate) to be shipped to various destinations serviced by the railroad, and collect units brought in by the railroad for them. This process is resource-intensive in the sense that it involves chassis, parking lots, lifting equipment, hostlers, hub personnel, crew, rail equipment, and so on. Units trickle in as customers bring in, and railroad needs to aggregate and assign them to an outbound train that can deliver them on time at the intended destination. Similarly, an inbound train brings in a large batch and railroad operations need to deramp (i.e. unload) and store the units in right parking spot in the hub before customer could pick them up, without hampering the fluidity of the hub facility.

This initiative employs two time-space networks – one for the in-gate traffic and the other inbound- to represent hub operations and process events as nodes, and unit flow between them as edges. The flow of units through both the networks is governed by resource availability and complementary and competitiveness between them. This dual-stream, time-space network model synchronizes the unit flow between ingate and inbound streams and facilitates resource exchange and allocation, and identifies scenarios where additional resources are needed from external sources to address the ebbs and flows of intermodal traffic through the hub facilities.

4 - Resource and capital planning for railroad automotive facilities using simulation

Andrea Arias, BNSF Railway, Fort Worth, TX, United States, Rajon Pantha, Justin Taylor

Efficient resource and capital planning are critical for railroad automotive facilities to maintain operational excellence and meet customer demands. In this presentation, we show how simulation can be leveraged to make data-driven decisions, optimize resource utilization, and strategically plan capital investments to enhance operational efficiency, customer satisfaction, and financial performance.

5 - A Dynamic Programming Algorithm for Intermodal Load Planning

Wenjie Li, BNSF, Fort Worth, TX, United States, April Kuo

In load planning, containers and trailers (i.e. units) from intermodal facilities are assigned to railcars on tracks with many operational considerations, such as characteristics of units, and other load compliance rules. Each unit has its service priority. Basically, units with low service priority cannot be assigned if units with high service priority have not been assigned. Thus, this load planning problem aims to satisfy service level agreements and follow compliance rules. A dynamic programming algorithm is designed to find optimal solutions in intermodal load planning.

MA40

Summit - 432

Empirical Research in Healthcare Operations

Invited Session

Behavioral Operations Management

Chair: Bing Bai, McGill University, Montreal, QC, Canada

Co-Chair: Nicholas Clyde, Washington University in St. Louis, St. Louis, MO, United States

1 - The Impact of Ridesharing Platforms on Healthcare Access

Nicholas Clyde, Washington University in St. Louis, St. Louis, MO, United States, Bing Bai, Dennis Zhang

The vast majority of people in the United States do not access the recommended healthcare services, and missed opportunities for prevention are costing the country an estimated $55 billion annually. This is particularly important for older patients, who face a higher risk of adverse health outcomes if potential issues are not detected early. Unfortunately, this population is also prone to transportation barriers because of mobility issues and driving limitations. Our paper uses a natural experiment to empirically investigate the impact of ridesharing platforms on healthcare access. We specifically focus on Medicare patients since healthcare access for them has the most potential to improve health outcomes and reduce spending for this population. We employ a difference-in-differences specification to show that ridesharing platforms have a negative impact on healthcare access for Medicare patients. We further show that this negative impact is mitigated by access to public transportation. Finally, we show evidence for two potential mechanisms: the re-entry of ridesharing platforms causes an increase in traffic and a negative impact on non-profit organizations which provide transportation services to older people and people with disabilities. Our results suggest that ridesharing platforms increase the inconvenience and financial cost of going to healthcare appointments.

2 - Contractors in Nursing Homes: How Contractors Influence the Quality of Care

Minyu Ye, University of Illinois Urbana-Champaign, Champaign, IL, United States

Contractors have become a ubiquitous presence in the operations of nursing homes, playing a pivotal role in shaping the delivery of care to residents. To explore the intricate relationship between the utilization of contractors and the quality of care provided in nursing homes, we employ an Instrumental Variable (IV) method to uncover the causality between the two and examine the underlying mechanisms driving these effects. Understanding the nuanced interplay between contractor usage and care quality is essential for enhancing the overall standards of care in nursing homes and ensuring optimal outcomes for residents. By utilizing the public holiday as the instrumental variable that are exogenous to the quality of care but influence contractor usage, we find that with higher usage of contractors in nursing homes, there are more residents experiencing accidental falling. However, there is no significant effect on Urinary Tract Infection (UTI). Furthermore, we find that the usage of contractors may lead to inconsistent working schedule of staff, which harms the familiarity between nurses and residents, eventually causing the worse quality of care in nursing homes.

3 - Does Team Familiarity Harm the Care Delivery Quality? Evidence from Cardiac Interventional Procedures

Xiaohan Chen, Xiamen University, Xiamen, China, People's Republic of

Managers are often advised to strive for high team familiarity to increase team coordination and thus improve team performance. However, long-term partnerships may reduce learning effectiveness. To explore the potential negative effects of team familiarity on the quality and efficiency of teamwork, we use data from emergency percutaneous coronary intervention (PCI) procedures to estimate the impact of team familiarity on the incidence of major adverse cardiovascular events (MACE) and the duration of each procedural stage. We find U-shaped effects of familiarity both on cooperative quality and efficiency, and these phenomena vary with the requirements of coordination in different task stages. We observe that cross-hierarchical relationships are the source of these negative effects. Finally, we test the moderating effect of leader authority on the negative effect of familiarity. We contribute to adding the evidence and mechanisms by which familiarity negatively affects team performance, particularly in the healthcare field. We discuss that the application of collaborative strategies lies in a dynamic focus on familiarity rather than always strengthening partnerships.

4 - All Horn and No Halo: Explaining Safety-Net Hospitals Underperformance on Patient Experience

Ankita Shirahatti, Boston University, Boston, MA, United States, Anita Carson

Safety-net hospitals play a critical role in the American healthcare ecosystem but face chronic financial challenges - especially on pay-for-performance programs like HCAHPS. Our study takes advantage of the expanse of publicly available information on hospital quality and financial performance to assess how accurately patient-satisfaction is measured for safety-net hospitals. We first examine whether a horn effect --- the opposite of a well-documented cognitive bias known as the halo effect --- distorts patient perceptions of safety-net hospital experiential quality. We also use the growing volume of online reviews of Massachusetts hospitals to identify dimensions of safety-net care that, while deeply important to their patients, are not included as patient experience measures on HCAHPS. We find evidence that safety-net hospitals are subject to a horn effect in which they lose up to two points on the best possible scores for patients' overall ratings and willingness to recommend their hospital. Our results are supported by an alternate empirical specification that examines the effect of newly assigned safety-net status on HCAHPS performance. Our textual analysis of online hospital reviews reveals that ratings are dependent on several features of care not measured by HCAHPS, including a hospital's wait times, diagnostic testing, accommodations, billing process, staff communication, and staff attention levels.

MA41

Summit - 433

Machine Learning for Queueing Model Discovery and Control

Invited Session

Applied Probability Society

Chair: Vahid Sarhangian, University of Toronto, Toronto, ON, Canada

Co-Chair: Berk Gorgulu, McMaster University, Hamilton, ON, Canada

1 - Admission Decisions Under Imperfect Classification: An Application in Criminal Justice

Zhiqiang Zhang, The University of Chicago Booth School of Business, Chicago, IL, United States, Pengyi Shi, Amy Ward

We study the decision of admitting individuals eligible for probation to incarceration diversion treatment programs, whose aim is to reduce the chance an individual re-offends. The admission decisions are made based on the outcome of a machine learning (ML) algorithm. We model the diversion program as a queueing loss model with reneging during service, and calibrate the model using data from Illinois. Using fluid optimization, we show that the optimal policy has a simple prioritization scheme when the ML algorithm's prediction is sufficiently accurate, but is suboptimal otherwise. We provide a finite sample bound for an imperfect prediction in tabular mixture models and propose guidelines for potential interventions such as collecting more data and involving human judgement. Finally, we investigate the fairness issues among different subpopulations. We explore various fairness criteria from ML and operations perspectives, and discuss the impact of each fairness criterion on admission decisions and corresponding outcomes.

2 - Data-Driven Stochastic Modeling Using Autoregressive Sequence Models: Translating Event Tables to Queueing Dynamics

Daksh Mittal, Columbia University, New York, NY, United States, Shunri Zheng, Jing Dong, Hongseok Namkoong

While powerful, stochastic modeling of service systems demands significant human resources and extensive expertise. To scale and democratize these benefits, we propose a new data-driven paradigm based on prediction models, leveraging recent advances in autoregressive sequence modeling in AI such as transformers. We utilize the rich operational data typically stored in the form of event tables, which record sequences of event types and inter-event times. We propose a meta-learning framework that leverages autoregressive sequence models trained on historical event tables to learn the underlying dynamics of the system. Our main theoretical insight justifies how next-token prediction enables stochastic modeling through autoregressive generation, by drawing connections between partial exchangeability of Markov chains and De Finetti's predictive view of Bayesian inference. As a proof of concept, we substantiate our approach on event tables generated from queueing networks, and demonstrate its benefits in simulation and uncertainty quantification. Our proposed framework lays the groundwork for scaling the impacts of stochastic modeling by leveraging engineering advances in AI and ease of training prediction models.

3 - Approximate Dynamic Programming for Multiclass Scheduling Under Slowdown

Berk Gorgulu, McMaster University, Hamilton, ON, Canada, Jing Dong, Vahid Sarhangian

In many service systems, service times of customers can be correlated with waiting times. Scheduling under such dependency is challenging as a Markovian state description requires keeping track of all customers' waiting history. We propose an approximate dynamic programming algorithm for multi-class scheduling with wait-dependent service times. Our algorithm can generate policies with simple structures and achieve strong performance which we illustrate in a healthcare setting using real data.

4 - Scalable Learning in Weakly Coupled Markov Decision Processes

Chen Yan, The University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Weina Wang, Lei Ying

We explore a general model of Markov decision processes (MDPs) consisting of a large number N of independent sub-MDPs, linked by global constraints. The complexity of finding optimal policies is known to grow exponentially with the number of components N.

In the non-learning scenario, when the model satisfies a specific non-degenerate condition, efficient (i.e., polynomial in N) Linear Programming (LP)-based algorithms exist, achieving a performance gap smaller than $\sqrt{N}$ relative to the LP upper bound value. Conversly, without this non-degenerate condition, any existing heuristics results in a gap of $\sqrt{N}$, and this Central Limit Theorem-type order is generally tight.

In this talk we will improve our current understanding of the order in terms of N for these bounds and then discuss the implications in the learning scenario, considering both the degenerate and non-degenerate settings.

MA42

Summit - 434

Large-Scale Stochastic Systems

Invited Session

Applied Probability Society

Co-Chair: Debankur Mukherjee, Georgia Tech, ISyE, Marietta, GA, United States

1 - Mean-Field Approximations for Stochastic Population Processes with Heterogeneous Interactions

Anirudh Sridhar, MIT, Cambridge, MA, United States, Soummya Kar

This work studies a general class of stochastic population processes in which agents interact with one another over a network. Agents update their behaviors in a random and decentralized manner according to a policy that depends only on the agent's current state and an estimate of

the macroscopic population state, given by a weighted average of the neighboring states. When the number of agents is large and the network is a complete graph (has all-to-all information access), the macroscopic behavior of the population can be well-approximated by a set of deterministic differential equations (the classical mean-field approximation). For incomplete networks such characterizations remained previously unclear, i.e., in general whether a suitable mean-field approximation exists for the macroscopic behavior of the population. The present work addresses this gap by establishing a generic theory describing when various mean-field approximations are accurate for \emph{arbitrary} interaction structures.
Our results are threefold. Letting W be the matrix describing agent interactions, we first show that a simple mean-field approximation that incorrectly assumes a homogeneous interaction structure is accurate provided W has a large spectral gap. Second, we show that a more complex mean-field approximation which takes into account agent interactions is accurate as long as the Frobenius norm of W is small. Finally, we compare the predictions of the two mean-field approximations through simulations, highlighting cases where using mean-field approximations that assume a homogeneous interaction structure can lead to inaccurate qualitative and quantitative predictions.

2 - Invariance Principles and Mckean-Vlasov Limits for Randomized Load Balancing In Heavy Traffic

Rami Atar, Technion, Haifa, Israel, Gershon Wolansky

A load balancing model is considered where service time distribution possesses a finite second moment (otherwise general). A small fraction of arrivals pass through the power-of-choice algorithm while the remaining ones are assigned to queues chosen uniformly at random. The system is analyzed at the many-server limit in heavy traffic. It is proved that the hydrodynamic limit exists and is characterized as the unique solution of a parabolic PDE with nonlocal coefficients. The PDE has an explicit stationary solution.

Next, two forms of an invariance principle are proved, one under a rather general initial distribution and another under exchangeability. The latter is given by a McKean-Vlasov SDE, a result closely related to limit results for Brownian particles interacting through their rank. However, an entirely different set of tools is required because of the non-Markovian nature of the queue length processes.

3 - A Large-Scale Particle System with Self-Propulsion and Distributed Synchronization

Alexander Stolyar, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Yuliy Baryshnikov

A system of n particles, moving forward in jumps on the real line, is considered. Each particle can make both i.i.d.-sized "independent" jumps and "synchronization" jumps, allowing it to join a randomly chosen other particle if the latter happens to be ahead. System state is the empirical distribution of particle locations. The mean-field asymptotic regime, where n becomes large, is considered. Our results include: (a) convergence of the process to the mean-field limit (MFL); (b) MFL propagation speed (depending on the initial state) and existence of traveling wave solutions; (c) proof of a "minimum speed selection principle," stating that the limit of steady-state propagation speeds is the minimum possible speed of a traveling wave MFL.

4 - Characterizing Rare Events for Interacting Markov chains on Sparse Graphs

Kavita Ramanan, Brown University, Providence, RI, United States, Sarath Yasodharan, I-Hsun Chen, Ivan Lee

Many phenomena in operations research and engineering can be modeled as interacting processes whose instantaneous evolution is governed by an underlying interaction graph. While there has been a lot of attention devoted to the case when the underlying graph is dense, in many real-world applications this underlying interaction graph is sparse. We develop a framework for describing probabilities of rare events, or large deviations probabilities, in such systems on sparse graphs. We also illustrate, via specific examples, how this framework can be applied to gain insight into how rare events occur, and how they are influenced by the underlying graph topology. The talk is based on various joint works with I-Hsun Chen, Ivan Lee and Sarath Yasodharan.

MA43

Summit - 435

Multiserver & Preemptive Scheduling

Invited Session

Applied Probability Society

Chair: Izzy Grosof, Northwestern University, Evanston, IL

1 - New Lower Bounds on Optimal M/G/k Scheduling

Isaac Grosof, Northwestern University, Chicago, IL, United States, Ziyuan Wang

A well-designed scheduling policy can significantly improve the performance of a queueing system, without requiring any additional resources. While scheduling is well-understood in the single-server setting, much less is known in the multiserver setting. Results are particularly sparse in moderate load settings, outside of the asymptotic regimes of heavy traffic and light traffic. Lower bounds are particularly sparse outside of these settings, limited to the M/G/1/SRPT and M/G/infinity lower bounds, which are only tight in these asymptotic limits.
We seek to prove tighter lower bounds on mean response time in the M/G/k, especially under moderate load. We introduce the WINE lower-bounding framework, allowing multiple lower bounds to be combined into a single, stronger lower bound. Moreover, we introduce and analyze the Increasing Speed Queue, which captures the variable-service-speed nature of the M/G/k system, and use it to further strengthen our lower bound on M/G/k scheduling.

2 - Simple Policies for Multiresource Job Scheduling

Zhongrui Chen, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Isaac Grosof, Benjamin Berg

Workloads in data centers are composed of multiresource jobs requesting a variety of computational resources including CPU cores, memory, disk space, and hardware accelerators. In general, a data center server can run multiple jobs in parallel, but a set of jobs can only run in parallel if the server has sufficient resources to satisfy the demands of each job. It is generally hard to find sets of jobs that perfectly utilize all server resources, and choosing the wrong set of jobs can lead to underutilization of the server. This raises the question of how to allocate resources across a stream of arriving multiresource jobs to minimize the mean response time across jobs — the mean time from when a job arrives to the system until it is complete.

Current policies for scheduling multiresource jobs are complex to analyze and hard to implement. We propose a class of simple policies, called Markovian Service Rate (MSR) policies. We show that the class of MSR policies is throughput-optimal, in that if a policy exists that can stabilize the system, then an MSR policy exists that stabilizes the system. We derive bounds on the mean response time under an MSR policy, and show how our bounds can be used to choose an MSR policy that minimizes mean response time.

3 - Transform Analysis of Preemption Overhead in the M/G/1

Shefali Ramakrishna, Cornell University, Ithaca, NY, United States, Ziv Scully

We present the first transform analysis of the M/G/1 queue under preemptive priority scheduling with preemption overhead. Preemptive scheduling allows high-priority jobs to interrupt low-priority ones, but such interruptions involve overhead costs that are often overlooked in queueing literature. Our work incorporates these overheads, accounting for stochastic delay each time a job is paused or resumed. We also analyze preemptive policies that could minimize overhead by making jobs "resistant" to preemption. Such adjustments could potentially balance the costs and benefits of preemption.

Along the way, our analysis develops a number of new tools for studying preemption overhead in queues. For example, one of the main technical challenges is that a low-priority job's service time is dependent on how many high-priority arrivals occur during it. We develop a method based on multivariable Laplace transforms to capture this dependence. The tools we develop are flexible enough to be useful beyond the specific model of preemption overhead we consider.

This research lays the groundwork for further theoretical exploration into preemptive scheduling with overhead, encouraging subsequent studies to incorporate overhead considerations into more complex scheduling policies.

4 - Transform methods for Markov-modulated queues

Daniela Hurtado-Lange, Kellogg, Northwestern University, Evanston, IL, United States, Isaac Grosof

The Transform Method based on the Drift Method has been developed in the last four years for heavy-traffic analysis in a variety of queueing systems, including the single-server queue, load-balancing system, input-queued switch, generalized switch under complete resource pooling, and ride-sharing. These are studied in discrete time under i.i.d. arrival and potential service process. In this talk, I will introduce a generalization to the Transform Method, where we allow a Markov-modulated arrival and service process. The key is in carefully designing the test function, which includes the queue length and a correction due to the Markov-modulated nature of the arrival process. In this talk, I will present a simple case, but the methodology can be generalized to more complex queueing networks.

5 - Heavy-traffic Optimal Known Size Scheduling in Generalized Switches

Runhan Xie, University of California, Berkeley, Berkeley, CA, United States, Isaac Grosof, Ziv Scully

We study the problem of known-size scheduling in generalized switches. The scheduler has available the size and class information of each job in the system and chooses which jobs to enter service. We give the first scheduling policy that provably minimizes mean response time in heavy traffic. Our results apply to two important classes of queueing models: multiserver-job systems and queues with compatibilities. We remove the server need restrictions, such as power-of-two, in prior study of multiserver-job scheduling and offer insights into scheduling in queues with compatibilities. Our result is a novel combination of ideas from load-balancing guardrails, approximate WINE, and MaxWeight policy.

MA44

Summit - 436

Simulation and Generative AI

Invited Session

Simulation Society

Chair: Zeyu Zheng, University of California, Berkeley, Berkeley, CA, United States

1 - Contractive Diffusion Probabilistic Models

Hanyang Zhao, Columbia University, New York, NY, United States, Wenpin Tang

Diffusion probabilistic models (DPMs) have emerged as a promising technology in generative modeling. The success of DPMs relies on two ingredients: time reversal of Markov diffusion processes and score matching. Most existing works implicitly assume that score matching is close to perfect, while this assumption is questionable. In view of possibly unguaranteed score matching, we propose a new criterion -- the contraction of backward sampling in the design of DPMs, leading to a novel class of contractive DPMs (CDPMs). Our key insight lies on the illustration that the contraction in the backward process can narrow score matching errors as well as discretization errors, thus our proposed CDPMs are theorectically robust to both sources of error. For practical concerns, we further show that CDPM does not require any retraining and can leverage pretrained existing DPMs by a simple transformation. We corroborated our proposal by experiments on synthetic 1-dim experiments, Swiss Roll, MNIST and CIFAR-10 dataset: notably, CDPM shows the best performance among all known SDE-based DPMs in terms of W_2 error and smallest FIDs for CIFAR10 unconditional generation, surpassing VE SDE based on NCSN++ \cite{Song20} architecture and the same predictor-corrector sampler.

2 - Learning to Generate Heavy-tailed Conditional Distribution via Diffusion Model

Tingyu Zhu, University of California, Berkeley, Berkeley, CA, United States, Haoyu Liu, Zeyu Zheng, Jinghai He

Emerging applications of conditional generative models in the domain of service queueing systems, supply chain, healthcare and revenue management often witness data with heavy-tailed distribution. Despite the importance of adequately learning heavy-tailed distributions for these applications, we theoretically show that existing conditional generative models, such as diffusion models, encounter challenges in addressing the tail distribution in both model training and data generation. To address these challenges, we develop a method to effectively capture the heavy-tailed distribution patterns, and propose a new framework for conditional generation of heavy-tailed data. We further design synthetic experiments to efficiently evaluate generative models for conditional distributions. Via the experiments, we demonstrate that our method shows superior performance in learning and generating heavy-tailed distributions. Our code is provided in supplementary materials.

3 - Supply Chain Financing Risk Management through a Quantile-regression-based Generative Model

Qingkai Zhang, City University of Hong Kong, Hong Kong, China, People's Republic of, Jeff Hong, Houmin Yan

Supply chain finance is an effective solution for addressing the financing challenges encountered by small and medium cross-border e-commerce enterprises. This study focuses on the crucial risk management issue of supply chain finance. In contrast to conventional approaches, this paper proposes an innovative application of generative models to effectively manage the risk associated with supply chain finance. Specifically, we propose employing a quantile-regression-based generative model to learn the probability distribution of e-commerce product sales, especially the left tail of the distribution which plays a crucial role in risk management. Once the generative model is constructed, any interested risk measure can be easily calculated. Building upon the unique characteristics of supply chain issues, we also derive the convergence and convergence rate of the generative model and corresponding loss of the loan under linear quantile regression assumptions. Moreover, in order to capture non-linear relationships and complex combination characteristics between sales and covariates, we enhance the linear model using deep neural network and factorization machine techniques. Numerical experiments conducted on synthetic and real-world data validate that our proposed generative model can accurately predict the target distribution and facilitate subsequent risk management.

4 - A class of grey-box diffusion models

Haoyu Liu, University of California, Berkeley, Berkeley, CA, United States, Tingyu Zhu, Zeyu Zheng

Diffusion models, which aims at learning a diffusion process that generates a probability distribution for a given dataset from random noises, have been shown to be successful in generation tasks of many areas, including but not limited to audio, graphs, text, time series, and general tabular data. The generation process starts from random noises from a known distribution (usually Gaussian distribution), and iteratively denoise until we obtain new samples from the target distribution. It requires a neural network based model that takes noisy sample as input and predicts the noise. For different kind of tasks, properly designed architecture of the noise prediction model is crucial for the success of the method. Commonly used architectures include U-Net and transformer-like blocks. However, most of the current networks are still black-box that are evaluated only on end-to-end performance. In this work, we propose some preliminary methods to obtain grey-box diffusion models, whose architecture has higher interpretability and allows for more control in the generation process.

MA45

Summit - 437

Economics of Pharmaceutical Markets

Invited Session

Health Applications Society

Chair: Qinquan Cui, UCL School of Management, University College London, London, United Kingdom

Co-Chair: Kenan Arifoglu, UCL School of Management, University College London, London, E14 5AA, United Kingdom

1 - The Role of Partnerships in Clinical Trials

Feryal Erhun, University of Cambridge, Cambridge, United Kingdom, Lidia Betcheva Haris, Houyuan Jiang, Efthymia Kostaki, Nektarios Oraiopoulos

Drug discovery and development is a lengthy and costly process which, on average, takes 10 to 15 years and costs $2.6 billion per drug, with the majority of R&D costs resulting from clinical trials. Meanwhile, the cost of running a trial has been increasing annually by at least 10%. The importance of launching new drugs to the market quickly, exacerbated by climbing R&D costs, has prompted pharmaceutical companies to strive to conduct clinical development in a more efficient manner. One direct way a pharmaceutical company can extract more value from a drug is by reducing the trial duration. In this talk, we explore how pharmaceutical companies do that with different models.

2 - Mitigating the U.S. Drug Shortages: Evaluating Quality Reward and Other Interventions

Hui Zhao, The Pennsylvania State University, University Park, University Park, PA, United States, Sergey Naumov, In Joon Noh

Drug shortages have been persistent in the U.S. for over a decade, posing serious threats to public health and the healthcare system. While previous research has investigated the causes and effects of drug shortages, there is a dearth of research exploring potential solutions to mitigate this problem. Using a system dynamics model of the U.S. generic drug market, we evaluate the long-term effectiveness of two existing policy interventions (expediting drug approvals and nudging manufacturers to ramp-up their production) and the "quality reward" initiative that is being actively explored by the FDA and industry. Our results indicate that while the effect of the existing policy interventions is limited, quality reward can mitigate drug shortages in a sustainable way. However, a potential caveat of quality reward is the emergence of a monopolistic supply market with negative consequences. We suggest that a carefully designed quality disclosure mechanism can address *this issue. To the best of our knowledge, this is the first study to quantitatively and comparatively evaluate long-term effectiveness of quality reward and other interventions on drug shortages and provide structural explanations for their performance.*

3 - The Role of Information Asymmetry Between Brand-Name and Generic Firms in Pharmaceutical Markets

Qinquan Cui, UCL School of Management, University College London, London, United Kingdom, Kenan Arifoglu, Dongyuan Zhan

Brand-name firms, leveraging their extensive resources and patent protection, possess more information about consumers compared to generic firms. This *information asymmetry* plays a crucial role in shaping pharmaceutical markets. Moreover, following the entry of their generic alternatives, some brand-name drug prices experience a paradoxical increase (i.e., the *Generic Competition Paradox (GCP)*). However, existing research overlooks the information asymmetry between brand-name and generic firms, failing to provide a comprehensive understanding of pharmaceutical markets. To bridge this gap, we develop a game-theoretic model with signaling to analyze interactions between a brand-name firm and a generic firm over two periods (signaling and full-information). We find that the brand-name firm can use *limit pricing* during the signaling period to deter generic entry by charging below its monopoly price, then increase its price and allow entry in the full-information period, leading to the GCP. Consequently, under information asymmetry, limit pricing may arise from the brand-name firm's *anticompetitive* practices, offering an *alternative* explanation for the GCP. While typically resulting in higher drug prices and reduced consumer choice, limit pricing and the GCP can sometimes *benefit* consumers and society. This challenges the prevailing belief that information asymmetry and the GCP are always detrimental, highlighting their nuanced impacts.

4 - Surprise! The Relative Effectiveness of Preannounced and Unannounced Inspections

Zach Wright, Ohio State University Fisher College of Business, Columbus, OH, United States, John Gray, In Joon Noh, George Ball

FDA is the sentinel for ensuring the quality of pharmaceuticals sold in the U.S. For decades, FDA has utilized its facility inspections program to perform periodic process inspections of drug manufacturers. Such inspections in the U.S. are nearly always unannounced, while inspections in foreign countries are preannounced, up to three months in advance. This study examines whether this variation in FDA inspection strategy may impact inspection and product quality outcomes. We employ FDA inspection data, which include two unannounced-inspection pilots performed in India. Using a coarsened exact matching (CEM) matching approach designed to mimic the FDA's inspection Site Selection Model, we examine both inspection outcomes and post-inspection serious adverse events for drugs made at the plant. We find no difference between the likelihood of the most extreme inspection outcomes occurring when comparing all inspections. However, when comparing unannounced inspections in India to pre-announced inspections in the US, we find that inspections in India are several times more likely to receive the egregious (e.g., poor quality) inspection outcomes, suggesting that preannouncing inspections may incentivize "window dressing behavior" by foreign pharmaceutical manufacturers. Further, we find evidence of decreases in serious adverse events following extreme inspection outcomes in India, pointing to the importance of accurate assessments of foreign drug facilities in inspections. Our findings highlight the importance of using unannounced inspections to improve drug outcomes.

MA46

Summit - 438

AI in Healthcare

Invited Session

Health Applications Society

Chair: Mehmet Ayvaci, The University of Texas at Dallas, Richardson, TX, 75287, United States

1 - What People Think of Machines as Doctors: Unveiling the Value of Gen-AI for E-Health

Dicle Yagmur Ozdemir, Erasmus University, Rotterdam, Netherlands, Mehmet Ayvaci, Alejandro Zentner

Large language models (LLMs) generate human-like text from vast data, enabling natural language communication even for tasks demanding expert knowledge. As LLMs increasingly become an alternative for experts, understanding how non-experts perceive and respond to automated responses by machines (in this case, LLMs) is crucial. Framed within the context of patient-physician communication, we investigate how non-experts (typical patients) perceive LLM responses versus physician responses and explore the factors influencing their perception. In a survey-based experiment, we compare non-experts' (survey participants) evaluations of responses from physicians and ChatGPT, a Chat Generative Pretrained Transformer, to patient queries. Our findings reveal that non-experts overwhelmingly prefer ChatGPT responses over responses by physicians, even when machine responses are of low quality (as judged by a blinded panel of experts). Two key factors influencing this preference emerge from our study: longer prose from ChatGPT heightens non-experts' preference for machines, while disclosing the response source diminishes this preference, especially when the ChatGPT response quality is lower. Our study indicates the need for a careful use of LLMs when responding to laypersons, particularly patients, in their search for answers to health-related questions.

2 - AI Orientation of Online Influencers: Evidence from A Fitness Platform

Fanruo Wang, University of Rochester, Rochester, NY, United States, Hongfei Li, Weiguang Wang

The influencer economy has flourished over the past decade, driven by social media marketing, endorsements, and product placements. While it could be cooperative or competitive among influencers, artificial intelligence (AI) is revolutionizing the landscape of the influencer economy. On one hand, AI can be cooperative with influencers by streamlining and enhancing content creation, spanning from idea generation to content writing. On the other hand, AI has the potential to challenge or replace influencers by producing high-quality content at scale and delivering timely, personalized responses. With the unprecedented change in the influencer economy, however, influencers' attitudes towards AI are unexplored. In this study, we conduct a large-scale field experiment by sending a fitness-related question in a private message to the influencers that contains the different activation information of value orientation towards human and AI peers and analyze their responses. Our findings indicate that influencers exhibit a greater inclination to respond when prompted with the activation of social value orientation, whereas they display reduced responsiveness when primed with the activation of AI value orientation. We further explore the heterogeneous attitudes towards AI and observe that influencers are less willing to contribute with the AI cooperative activation.

3 - Technological Interventions to Combat Social Stigma in Mental Health Support

Zhichen Chen, Boston University, Boston, MA, United States, Gordon Burtch, Bin Gu, Chrysanthos Dellarocas

Many individuals in the United States face mental health challenges but do not seek professional help. Besides cost and accessibility, stigma is a significant barrier. This study explores how telehealth and conversational artificial intelligence (AI) can reduce stigma and encourage individuals to seek mental health support. We conducted a controlled online experiment with over 1,000 participants who reported mental health challenges. Participants were randomly assigned to two independent treatments: one manipulated their expectations of stigma, employing a social norm intervention, while the other presented them with alternative bundles of mental health service offerings (combinations of offline services, online telehealth services, and a ChatGPT-like conversational AI agent with varying data privacy assurances). Results showed that participants had greater intention to engage with mental health services when telehealth and conversational AI options were provided, especially when they were treated with high stigma concerns. However, when treated with low stigma concerns, participants' intention to engage with offerings involving conversational AI depended on strong data privacy commitments. Weak privacy assurances reduced interest in using offerings involving conversational AI. These findings highlight the potential of telehealth and conversational AI to increase engagement with mental health services. They also underscore the need for strong data privacy policies to address the tension between social stigma and privacy concerns. Implications for designing and implementing conversational AI for mental health support are discussed, considering data privacy and demographic factors.

4 - Between Learning and Practicing: Investigating Labeling Availability related AI Fairness in Predicting Diabetes Progression from Prediabetes

Junjie Luo, Johns Hopkins School of Medicine, Baltimore, MD, United States, Ritu Agarwal, Gordon Gao

AI fairness is increasingly crucial as AI technologies are implemented in critical areas. Current research often presumes that data in practical settings are of full labeling availability as those in learning settings. This assumption can be flawed, as not all individuals from practical settings return for follow-up to generate the label to be learned by models. This discrepancy in feature and labeling distributions between learning and practicing settings is underexplored in the context of AI performance and fairness evaluations. Addressing this, we examine the impact of labeling availability on model fairness, using diabetes progression prediction from prediabetes as a case study. Our analysis reveals that only 37.09% of data points in practical settings are labeled for model training. Furthermore, there is a significant racial disparity in labeling availability: 40.33% for White individuals compared to 35.26% for Black individuals. This discrepancy not only affects dataset construction but also extends to model performance, where the prediction accuracy for Black individuals is 2% lower than for White individuals. To tackle these issues, we designed a model that evaluates labeling availability for each data point. Our findings show that data points with higher labeling availability scores generally yield better prediction performance, with accuracy improving by 0.26% for a 1% increase in labeling availability. Most importantly, controlling for labeling availability eliminates racial disparities in prediction performance, highlighting the necessity of considering labeling availability in both model training and application phases. This paper calls attention to labeling availability for more equitable AI outcomes.

MA47

Summit - 439

OR and Other Analytical Approaches for Healthcare Systems

Invited Session

Health Applications Society

Chair: Ankit Bansal, State University of New York, Binghamton, Binghamton, NY, United States

1 - Data-Driven Robust MDPs: an Application to Clinical Treatment Decisions

Angela Lin, Massachusetts Institute of Technology, Cambridge, MA, United States, Gavin Findlay, Dessislava Pachamanova, Georgia Perakis

We propose an approach for solving empirically-learned Markov Decision Processes (MDPs) which addresses two important goals in the context of clinical decision-making: robustness to estimation errors and risk-aversion. Since the optimal treatments (policies) are heavily dependent on the empirically-estimated transition probabilities of the MDP, we leverage statistical tools and observed data to construct appropriately conservative uncertainty sets for our robust MDPs and obtain robust treatment policies. These treatment policies not only protect against estimation errors but also provide a risk-aware decision-making lens. We test our robust/risk-aware MDP methodology on a sepsis treatment problem using data from a Boston area hospital.

2 - Equitable Vaccine Allocation through a Dynamic Epidemic-Supply Chain Model

Kimiya MohammadiJozani, Virginia Tech, Blacksburg, VA, United States, Esra Buyuktahtakin Toy

We present an optimization framework that integrates epidemic compartmental models with vaccine supply chains, incorporating equity considerations. Our model offers a proactive vaccination strategy that effectively controls disease spread by anticipating disease waves based on real data and planning mitigation efforts in advance. It serves as a dynamic decision-support tool for policymakers, optimizing vaccine distribution with various definitions of equity to ensure fairness. We introduce a new decomposition and cutting plane algorithm designed to address the computational complexities of our large-scale NP-Hard problem. This approach is demonstrated through its application to COVID-19 vaccine allocation in the US, showcasing its effectiveness in managing the complications of vaccine allocation challenges.

3 - Personalized Prediction and Control of Blood Glucose Levels using Dynamic System Modeling

Anjolaoluwa Popoola, Georgia Institute of Technology, Atlanta, GA, United States, Pooyan Kazemian, Kamran Paynabar

Patients with diabetes in the ICU typically rely on two methods to monitor their blood glucose levels: blood tests analyzed in the laboratory and fingerstick tests. While fingerstick tests offer a cost-effective and convenient approach, they sacrifice accuracy, providing only approximate measurements. In this paper, we propose a dynamic system model combining deep learning models and Kalman filtering to accurately predict blood glucose trajectory and assist with personalized glucose control over time. The resulting model, called a deep Kalman filter (Deep-KF) model, can be applied in the ICU and other healthcare settings. We validate our model using randomized real-world glucose and insulin clinical data. We demonstrate how our model can deliver precise blood glucose forecasts and insulin delivery recommendations and be used as a decision-support tool for healthcare providers, hence making glucose monitoring more efficient and patient-centric.

4 - *A Branch-and-Price Approach for Resource Allocation in a Pre-admissions Testing Clinic*

Mohammad Al Syouf, Binghamton University, Binghamton, NY, United States, Ankit Bansal, SALIGRAMA AGNIHOTHRI

Pre-admissions testing (PAT) clinics are hospital units serving pre-operative patients and performing tests to prepare patients for surgery. Patients may need multiple tests, each performed by a specialized nurse. Once a patient is taken to one of the limited exam rooms available, the patient stays in the room until all the required tests are completed. This research introduces a new modeling framework for optimizing the allocation of resources, such as room time and nurse time, in a PAT clinic. Efficient allocation of these resources is vital for patients, healthcare providers, and administrative personnel. The study proposes a novel routing-based modeling approach for resource allocation within PAT scheduling and presents a Branch-and-Price algorithm to address the problem. Computational results, based on real instances, are presented to demonstrate the effectiveness of the proposed approach.

5 - Minimizing Nurse Anesthetist Intraoperative Handovers

Abhishrut Sinha, State University of New York, Binghamton, Binghamton, NY, United States, Ankit Bansal, Osman Ozaltin, Michael Russell

Intraoperative handovers between Certified Registered Nurse Anesthetists (CRNAs) during surgeries pose a significant risk of medical errors, potentially leading to adverse health consequences. To address this, we introduce a two-stage stochastic optimization model that assigns CRNAs to operating rooms. The model aims to minimize intraoperative handovers while accounting for system operational constraints and uncertainties in surgery durations. We present computational results using data from a hospital system to showcase the effectiveness of our proposed approach.

MA48

Summit - 440

Topics in Platform Economy

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Dan Zhang, University of Colorado, Boulder, CO, United States

Co-Chair: Chengyi Lyu, University of Colorado Boulder, BOULDER, CO, 80309, United States

1 - Platform Entry and Vendor Competition in on-Demand Economy

Xiaotang Yang, University of Toronto, Toronro, ON, Canada, Arseniy Gorbushin, Ming Hu, Yun Zhou

We examine the impact of introducing a delivery platform on vendor competition. We consider an on-demand delivery system with n vendors competing for customers on the *full price*, which includes both the food price and the delivery fee. Vendors are strategic; each must decide whether to participate in the platform or build its dedicated delivery fleet (i.e., employing in-house delivery) to maximize its profit, or not offer any delivery option (i.e., only serving local customers). Vendors who participate in the platform set their food prices while the platform sets the delivery fee and takes a commission from the vendors' revenue. Vendors who build their own dedicated delivery fleets decide on both the food prices and delivery fees. To operate in the market, both the platform and vendors that build dedicated delivery fleets must ensure that their delivery times do not exceed the maximum allowable delivery time. We solve for the equilibrium of the underlying system and benchmark against the system without the platform. Our findings indicate that the introduction of the platform can either intensify vendor competition by subsidizing the delivery service (i.e., charging a delivery fee lower than the marginal delivery cost), leading to lower vendor profits and higher customer surplus, or conversely alleviates vendor competition by deriving profits from the delivery service, leading to higher vendor profits and lower customer surplus. Hence, the platform enabled sharing of couriers improves delivery efficiency, lowers delivery cost, but may intensify vendor competition and hurting vendors as in a prisoner's dilemma.

2 - Does Size Matter for Loyalty Points Redemptions?

Yang Chen, Smith School of Business, Queen's University, Kingston, ON, Canada, Anton Ovchinnikov, Nicole Robitaille

Prior research finds that redeeming loyalty points is generally associated with increased customer loyalty. However, does the size of a redemption impact loyalty? Numerous economic, behavioral, and psychological theories suggest that size might matter, but their specific predictions are often conflicting. Furthermore, whether a reward redemption feels small or large may be individual and context-dependent. With a large longitudinal proprietary dataset from a major loyalty program (LP) capturing its highest earners, our study investigates the effect of reward redemption size and consumer "type" on long-term behavioral loyalty. We find that size does matter in reward redemptions, and so does the operationalization of size and prior reward redemption behaviors. Small redemptions are sometimes associated with increased loyalty, whereas medium redemptions are most consistently associated with increased loyalty. With large redemptions, however, we sometimes observe a decrease in loyalty. Our findings suggest that encouraging the optimized redemption size has the potential to increase LP revenue significantly. That is, rather than passively letting reward redemptions happen, firms may want to actively manage them and treat redemptions as a new control lever in LP revenue optimization.

3 - Optimizing and Learning Assortment Decisions in the Presence of Platform Disengagement

Mika Sumida, Marshall School of Business, University of Southern California, Los Angeles, CA, United States, Angela Zhou

We consider a problem where customers repeatedly interact with a platform. During each interaction with the platform, the customer is shown an assortment of items and selects among these items according to a Multinomial Logit choice model. The probability that a customer interacts with the platform in the next period depends on the customer's past purchase history. The goal of the platform is to maximize the total revenue obtained from each customer over a finite time horizon. First, we study a non-learning version of the problem where consumer preferences and return probabilities are completely known. We formulate the problem as a dynamic program and prove structural properties of the optimal policy. Next, we provide a formulation in a contextual episodic reinforcement learning setting, where the parameters governing contextual consumer preferences and return probabilities are unknown and learned over multiple episodes. We develop an algorithm based on the principle of optimism under uncertainty for this problem and provide a regret bound. Previous approaches that address user disengagement often constrain exploration. However, in our model with non-permanent disengagement with assortments, the optimal solution simply offers larger assortments at the beginning of the horizon and exploration is unconstrained during the learning process. We numerically illustrate model insights and demonstrate regimes where our algorithm outperforms naively myopic learning algorithms.

4 - Customer Reward Programs for Two-Sided Markets

Chengyi Lyu, University of Colorado Boulder, BOULDER, CO, United States, Dan Zhang

Some two-sided markets have been spending billions of dollars on customer reward programs every year. Yet, there is little research on the rationale and impacts of such programs, despite the large literature on customer reward programs in traditional markets. This paper examines the rationale and impacts of customer reward programs in two-sided markets, investigating the efficacy of such programs and highlighting their interplay with matching schemes. We adopt an analytical model for a platform that interacts with customers and service providers over an infinite time horizon. Under a customer reward program, a customer earns a cash reward for every purchase with a finite expiration term, which can be used to offset the selling price in a subsequent purchase within the expiration window. Customers are heterogeneous in their request probabilities and valuations, while providers differ in their service costs. The platform maximizes its profit by setting the service compensation for providers, and the price and reward program parameters for customers. We show that adopting a customer reward program can often dramatically improve the platform's profit. Importantly, matching schemes play an important role -- customer reward programs are much more lucrative under priority matching schemes than a random matching scheme. Our results rationalize the heavy investments made by many two-sided markets in customer incentives and demand/supply matching. Overall, we conclude that customer reward programs can be an even more important profit-boosting tool for two-sided markets than for traditional one-sided markets. We also discuss the welfare implications of our findings.

MA49

Summit - 441

Applications of Stochastic Models in Service Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Demet Batur, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - Curbing the Opioid Crisis: Optimal Dynamic Policies for Preventive and Mitigating Interventions

Sina Ansari, Driehaus College of Business, DePaul University, Chicago, IL, United States, Shakiba Enayati, Raha Akhavan-Tabatabaei, Julie Kapp

This paper addresses the challenge of effectively responding to the opioid epidemic stemming from prescription pills through a public health lens. It centers on the strategic distribution of resources across diverse interventions aimed at preventing and mitigating the consequences of opioid use disorder (OUD) and overdose occurrences.

This paper proposes a decision aid tool built on the expected utility theory that leverages a Susceptible-Infected-Removed compartmental model to simulate the dynamics of the epidemic in a population. This model then feeds into a Markov Decision Process (MDP) model to generate optimal policies upon the current state of the epidemic. The optimal policies allocate the intervention budget to primary preventive and mitigating interventions in each decision period by minimizing the cost of fatal overdoses relative to the population's number of individuals with OUD, considering the impact magnitude of each intervention, based on the current state of the epidemic. A 10-year simulation of the epidemic's progression is conducted to assess the dynamic efficacy of the proposed decision tool.

The findings reveal an average reduction of 29% in total costs compared to the scenario without interventions and a decrease of 12% in total costs on average compared to the scenario with a 50-50 allocation. This study provides practitioners with a tool to effectively address the opioid epidemic and enhance public health by deciding how to allocate their budget to various levels of intervention.

2 - Dynamic Capacity Sharing in Wireless Networks

Demet Batur, University of Nebraska-Lincoln, Lincoln, NE, United States, Jennifer Ryan, Mehmet Vuran

A Markov Decision Process (MDP) model is proposed to study the decision problem faced by the operator of a private 5G network, known as a private cell, who must allocate available capacity to meet the resource needs of the primary user of the network, but who may also make excess capacity available to external secondary users to generate additional revenue. Private cells are privately owned wireless networks independent of commercial or public 5G networks. Industries such as manufacturing and transportation utilize private cells to prevent downtime for their automated operations. Private cells provide dedicated wireless connectivity, which enables enhanced security, reliability, and cost-effectiveness. They make use of network slicing, a technological advancement made available by 5G, to meet differentiated needs across applications. The network and its resources can be segmented through slicing to support specific applications. We study the problem faced by a private cell operator whose main responsibility is to serve the slice instances of the private cell's primary user but who can also lease excess resources to a secondary user to generate revenue. Slice instances require a slice with a specific combination of resources, such as spectrum, computation, or storage. The operator must manage the network's operations to determine which instances will be admitted to the network for service and which resources can be leased to the secondary user. We use an MDP model to formulate this problem and characterize the optimal admission and leasing decisions, which must be made in real-time based on the private cell's state.

3 - The Interplay Between Customer Feedback Solicitation and Product Innovation: a Dynamic Solution

Izak Duenyas, University of Michigan-Ann Arbor, Ann Arbor, MI, United States, Hanqi Wen, Joline Uichanco

Nowadays customers are becoming important partners of product/service innovation, with firms using customer feedback to guide product quality improvements. Soliciting customer feedback not only supports product innovation, it also cultivates loyalty among customers who feel that their voice can lead to change. However, soliciting informative feedback could be costly, leading to complicated tensions in the decision of whether to solicit feedback or not. In this paper, we analyze the dynamic policy for soliciting feedback that is informative for product quality improvement under two different settings where (i) the firm continually invests and (ii) the firm jointly decides when to invest in product quality improvement respectively. In setting (i), we find that the optimal policy has a monotone-threshold structure, which suggests the firm should initially engage in soliciting customer feedback, until either the amount of solicited feedback is sufficient or the product quality reaches a certain level, indicating the termination of solicitation. The policy is very easy to implement in practice. In setting (ii), the problem is extended to be the joint optimization of feedback solicitation and quality improvement, where we propose a simple but effective heuristic that only needs three thresholds. We run numerical experiments to verify the performance of the heuristic. Our numerical experiments reveal the importance of ensuring adequate customer engagement before committing resources to product innovation, where just a little coordination between feedback solicitation and product innovation can go a long way.

4 - Stability of Fork-Join Systems with Redundancy under Flexible Capacity Allocation

Chutong Gao, Northwestern University, Evanston, IL, United States, Seyed Iravani, Ohad Perry

We consider the stability problem of fork-join systems with redundancy under flexible capacity allocation. In this setting, a job consists of multiple tasks, each of which is sent to a different queue, and the service of a job is completed once some of its tasks are processed by the servers (so that the rest of the tasks are redundant). Further, servers can collaborate in processing tasks, thus increasing the rate at which they are processed. In the case when two tasks for each job must be completed, we establish the necessary and sufficient condition for systems under any static capacity allocation policy to be stable, and characterize the maximal stability region among all static policies. We then show that a static allocation policy achieves the maximal stability region if and only if the capacity allocated to each queue does not exceed one half of the total capacity. Next, we consider dynamic allocation policies, which allow for reallocation of capacity at any time. We show that dynamic policies achieve the same maximal stability region as the static policies, but under a milder condition, which only requires that the capacity allocated to the shortest queue does not exceed half of the total capacity at any time.

MA50

Summit - 442

Advances in Empirical Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: RJ Niewoehner, Kelley School of Business, Indiana University, Bloomington, IN, United States

Co-Chair: Maria Ibanez, Kellogg School of Management at Northwestern University, Evanston, IL, United States

1 - E-Access Versus Physical Access: An Examination of Telehealth Adoption

Eric Xu, Mississippi State University, Mississippi State, MS, United States, Kevin Linderman

As interest in telehealth has expanded in recent years, it is more important than ever to understand the tradeoffs patients make when deciding about whether to use in-person services or telehealth services. Therefore, we examine the impact of telemedicine adoption on primary care use, specifically examining the conditions by which telehealth consultations become a substitute for in-person care or a compliment to in-person care. Using a unique dataset of insurance claims for primary care visits, we examine in-person, asynchronous, and synchronous primary care visits.

2 - From Full-TIME to Flexi: Examining the Outcomes of Workforce Changes in Primary Healthcare Delivery

Harshita Kajaria-Montag, Indiana University, Bloomington, IN, United States, Michael Freeman

Primary care practices have responded to the workforce crisis by altering the
composition of their workforce, specifically, by hiring more part time and locum physicians. But, this is counterproductive as research suggests that providing continuity of care is productivity-enhancing, which is difficult to offer with such a fragmented workforce. Consequently, the trend suggests that continuity of care has been declining over the past decade. This paper examines this question from an operations management perspective by exploring the relative importance of workforce composition related operational factors that may explain variation in rates of COC between practices and over time.

3 - Machine Learning-Guided Cancer Screening: The Benefits of Proactive Care

Minje Park, The University of Hong Kong, Hong Kong, Hong Kong, Carri Chan, David Vawdrey, Elliot Mitchell, Abdul Tariq

With the advance of data analytics, many disease prediction models have been developed with the intent of detecting diseases earlier and improving patient outcomes. The operationalization of interventions and care based on these predictive models is critical to attaining these goals. We study the effects of a machine learning-guided colorectal cancer screening program. Using a regression discontinuity design based on the predicted risk score for having cancer, we find that the program increases the likelihood of colonoscopy uptake in three and six months by 6.8 percentage points (243% increase relative to the control sample within the bandwidth) and 9 percentage points (184% increase), respectively. Importantly, we also find significant effects on mortality. We estimate that the program decreases 2-year mortality by 7.2 percentage points (49% decrease). Our finding suggests that a machine learning-guided cancer screening program could significantly improve patient outcomes in addition to higher disease detection rates.

4 - The Impact of Batching Behavior in Emergency Department Prescription Orders

Yong Xia, University of California, San Diego, La Jolla, CA, United States, Danqi Luo

This study investigates the batching behavior of emergency department (ED) physicians in placing prescription orders and its impact on healthcare delivery quality. Batching behavior is defined as the practice of placing orders for multiple patients within a brief time interval immediately after seeing these patients. Understanding this behavior is crucial as it provides insights into workflow practices that influence healthcare outcomes and inform policy and operational improvements. Data from two comprehensive sources were analyzed: a large teaching hospital in Southern California over three years and a central hospital in China over seventeen months. The analysis reveals that physicians engaging in batching behavior order more tests, particularly lab tests, but fewer imaging tests. Furthermore, batching behavior is associated with a reduction in patient length of stay (LOS) but an increase in readmission rates. These findings suggest that while batching can improve workflow efficiency, it may adversely affect the quality of care. Thus, hospitals must consider the tradeoff between operational efficiency and maintaining high healthcare standards.

MA51

Summit - 443

NFT and Platform Innovation

Invited Session

MSOM: iForm

Chair: Yao Cui, Cornell University, Ithaca, NY, United States

Co-Chair: Jingchen Liu, Renmin University of China, Beijing, 100872

1 - Nft Market Design

Jingchen Liu, Renmin University of China, Beijing, China, People's Republic of, Yao Cui

In this paper, we study two NFT platforms that compete in a two-sided market over two periods. Period 1 corresponds to the primary sale period, where the creators mint and sell the NFTs to speculators on one of the two platforms. Period 2 corresponds to the resale period, where speculators resell the NFTs to collectors on the platforms. Under both centralized and decentralized royalty policies, we study the impact of platforms' adoption of interoperability on the competition between platforms and the welfare of other stakeholders.

2 - Unveiling the Scalping Factor: An Empirical Analysis of Nft Market Dynamics and Operational Insights

Wenbin Wang, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Shanshan Guo, Yue Qiu, Tian XIE

This paper explores the dynamic relationship between scalping activities and market returns in the non-fungible token (NFT) market. Central to our approach is the development of a theoretical model that elucidates the interaction between consumer and scalper behaviors, examining their impact on NFT pricing and market dynamics. Complementing the theoretical insights, our empirical analysis is anchored in a one-period-ahead regression model, utilizing the common correlated effect (CCE) estimation method. This method is crucial for identifying and quantifying an unobserved 'scalping factor', significantly influencing NFT returns. Analyzing the data on 16 Ethereum-based NFTs from OpenSea, we uncover a varying impact of scalping on various market segments.

Our research uncovers that while scalping negatively correlates with returns for larger-cap NFTs, it benefits smaller-cap NFTs, exhibiting immediate yet transitory effects in the dynamic NFT market. This research highlights the complexity of the NFT market and emphasizes the importance of the CCE methodology in unraveling the nuanced effects of scalping, offering key insights for investors and platform developers in this rapidly evolving sector.

3 - Information Good Pricing Under Unit Costs: Impact of Lazy Minting on Pricing Strategies In Nft Marketplaces

Giannis Kanellopoulos, Tilburg University, Tilburg, Netherlands, Dominik Gutt, Murat Tunc, Ting Li

Non-fungible tokens (NFTs) represent a new category of information goods, and unlike others, their marginal costs per additional unit are non-negligible. To alleviate the upfront financial costs typically borne by creators, NFT platforms introduced the "lazy minting" approach. Lazy minting introduces a new cost structure for information goods by deferring the minting fee until the point of sale and transferring the responsibility from the creators to the initial buyer. This shift has the potential to reshape the pricing dynamics of information goods, where production cost is no longer the primary factor guiding creators' pricing decisions. To explore this, we empirically examine the impact of lazy minting on NFT prices. Our findings reveal that creators' pricing decisions are influenced by their trading experience (i.e., the risk premium effect) and the duration they own and hold their NFTs (i.e., the endowment effect). Our results contribute to the pricing and valuation of information goods in response to the changes in production costs.

4 - When Crowdsourcing Meets Crowdfunding: The Work-Donation Interplay and Its Implications for User Engagement

Xue Tan, Southern Methodist University, Dallas, TX, United States, Rowena Gan, Kyunghee Lee

When platforms integrate crowdsourcing and crowdfunding, users play a dual role of both workers and donors. Understanding the interplay between individuals' work and donation behavior helps enhance user engagement on and contribution to the platform. We examine the interplay between users' decisions in a Web 3 community both analytically and empirically. Our empirical context was a digital platform (GitCoin) that offers both grants (a crowdfunding service) and bounties (a crowdsourcing service). While our economic model predicts that people will behave in a consistency-seeking way such that those who used to donate more would donate more in the future and those who used to earn more would work more in the future, our empirical analysis identified variety-seeking behavior in the short run. That is, workers who have recently earned incomes would be more likely to donate, and they would be most likely to donate when they have earned income from traditional projects. They would be least likely to donate when they earned the income from contests. In the long run, our prediction of consistency-seeking was validated in the data.

MA52

Summit - 444

Operational Decision-Making Under Uncertainty

Invited Session

MSOM: Service Operations

Chair: Milad Armaghan, Southern Methodist University, Dallas, United States

Co-Chair: Eojin Han, University of Notre Dame - Mendoza College of Business, Notre Dame, IN, United States

1 - Optimal Control of In-store Inventories for Omnichannel Retailing

Joohyun Kim, University of North Carolina at Charlotte, Charlotte, NC, United States, Sila Cetinkaya

We investigate the problem of optimizing the omnichannel inventories of a firm that owns and operates multiple retail stores in a stochastic dynamic setting. We consider the case where the relationship between online and in-store demands is considered explicitly and propose a new mean-field control model. We compute the closed-form solution characterizing the optimal control policy.

2 - Relative Monte Carlo for Reinforcement Learning

Audrey Bazerghi, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Sebastien Martin, Garrett van Ryzin

We introduce relative Monte Carlo (rMC), a new general purpose policy gradient algorithm for reinforcement learning with discrete action space. The policy is improved in real time using relative returns between a root sample path and counterfactual simulated paths instantiated by taking a different action from the root. The method is compatible with any differentiable policy, including the leading choice of neural network parametrization. It is guaranteed to converge for episodic and average reward tasks. Unlike traditional Monte Carlo, rMC policy gradient steps are performed throughout the rollout using a memory-efficient update decomposition, inspired by eligibility traces. Strong couplings between root and counterfactual paths further contribute to low data generation and memory requirements. We test rMC with a policy network in a two-tiered inventory fulfillment problem. Numerical results show it performs well compared to related algorithms.

3 - Predicting demand for wildfire suppression resources

Yasser Zeinali, University of Alberta, Edmonton, AB, Canada, ILBIN LEE, Mostafa Rezaei

Wildfires, as a global threat, pose significant damage recently in North America, where Canada reports over 8,000 wildfires annually, burning millions of hectares. The situation escalated in 2023, with Canada facing wildfires over an unprecedented 18.5 million hectares. Wildfire management agencies are under intense pressure to devise an effective resource allocation strategy. In this work, as a collaboration with the Alberta government, we integrate machine learning, queuing theory, and robust mixed integer programming to develop a predictive and prescriptive system for deploying wildfire suppression resources. This system estimates both the total daily resource-hours and the number of wildfires requiring each resource type, which together help us in determining the optimal number of resources to be deployed. Although our case study focuses on Alberta, our methodologies are adaptable and can potentially be tailored to meet the challenges of wildfire management in any region.

4 - A Competitive Newsvendor Problem under Limited Information

Milad Armaghan, Southern Methodist University, Dallas, TX, United States, Eojin Han, Sila Cetinkaya

We study a newsvendor game involving substitutable products. The industry demand is random and unmet demand for a retailer may spill over to competition. The retailers set their inventory levels with limited information about the industry demand distribution. Our goal is to characterize the distributionally robust inventory policies of the retailers. We discover the conditions under which it is optimal to disregard or consider the spill-over demand when setting optimal stocking levels. We also compare and contrast our problem setting to a monopoly under limited-information, and we investigate the conditions under which competition increases or decreases the industry inventory.

MA53

Summit - 445

Operational Data Analytics

Invited Session

MSOM: Supply Chain

Chair: Qi Feng, Purdue University, West Lafayette, IN, United States

Co-Chair: Lei Li, Hong Kong Polytechnic University, M507n, The Hong Kong Polytechnic University, Hong Kong, N/A

1 - Transfer Learning, Cross Learning and Co-Learning Across Newsvendor Systems with Operational Data Analytics (Oda)

Lei Li, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Qi Feng, J Shanthikumar

Decision making with limited statistical characterization and limited data is challenging. The typical statistical-machine-learning approaches would call for migrating the experience of a related system with ample data through transfer learning or leveraging the similarity of multiple systems with limited data through data pooling. We, instead, develop new solution concepts to learn across related systems by adapting the parametric Operational Data Analytics (ODA) framework for non-parametric decision making. We demonstrate that transfer learning can, indeed, improve decision performance in the focal system by utilizing a model pre-trained using ample data in a related system. In contrast, the ODA cross-learning approach utilizes the ample data from the related system to mimic the stochastic environment of the focal system. When the data from the old system are sufficiently large, the cross-learning solutions derived outperform any transfer-learning solution, and they are shown to asymptotically approach the parametric ODA solutions. When there are multiple related systems with limited data, we aggregate the data from different systems to create a generic stochastic environment for the decision-making problem. We show that the derived co-learning solutions are asymptotically optimal for the aggregate system and for each sub-system. This approach outperforms the existing data-pooling techniques in the sense that the latter focuses only on the aggregated performance, and the chosen solution may be (asymptotically) suboptimal for individual sub-systems. Our results underscore the roles of domain knowledge and the structural relationships between the data and the decision in designing efficient learning solutions with limited data.

2 - Conformal Contextual Optimization with a Smart Predict-Then-Optimize Method

Paul Grigas, UC Berkeley, Berkeley, CA, United States, Hyungki Im

We study an extension of contextual stochastic linear optimization (CSLO) that, in contrast to most of the existing literature, involves inequality constraints that depend on the uncertain parameters that are predicted by the machine learning (ML) model. Building on previous work that develops the “Smart Predict-then-Optimize (SPO)” loss and its tractable SPO+ surrogate loss in the case of a known deterministic feasible region, and work that develops robust variants of contextual optimization using conformal prediction methods, we propose a “Conformal Smart Predict-then-Optimize (CSPO)” approach for addressing uncertainty in the constraints. Specifically, we first propose the CSPO loss – a direct extension of the SPO loss – that measures the decision error, or regret, induced by following a robust predict-then-optimize approach that uses a conformal prediction method to produce an uncertainty set. We then propose a convex surrogate loss function called the CSPO+ loss – a direct extension of the SPO+ loss – to tractably train a prediction model in our CSLO setting. To train the model, we employ a data filtering procedure to address issues of infeasibility whenever the predicted uncertainty set fails to cover the true parameters. Furthermore, we utilize importance sampling when training to address the distribution shift induced by the data filtering step. Theoretically, we establish statistical consistency of the CSPO+ loss relative to the CSPO loss and provide finite-sample convergence guarantees for the CSPO loss with importance sampling under mild assumptions. Experimentally, we demonstrate strong performance of the CSPO+ loss on several different CSLO problem classes.

3 - A Single-Index Contextual Bandit Problem

Xi Zheng, University of Washington, Seattle, WA, United States, Gaoqian Xu, Wanning Chen

The Multi-Armed Bandit with Covariates (MABC) framework is a simple yet powerful approach for online decision-making under uncertainty. It has proven valuable in various fields, including online advertising, dynamic pricing and personalized medicine. Traditional MABC frameworks based on parametric and non-parametric models have been well studied. Parametric models are limited by structural assumptions and therefore lack flexibility, while non-parametric methods face significant challenges when the covariate dimension is high. Semi-parametric formulations offer an attractive middle ground between parametric and non-parametric estimation. In this paper, we consider a semi-parametric contextual bandit problem where the conditional expectation of the reward for each arm follows a single-index model. We propose two algorithms for this bandit problem and conduct a finite-time regret analysis. Additionally, we explore the inference properties for finite-dimensional parameters in the semi-parametric model. Numerical experiments based on synthesized and real-world datasets demonstrate that the practical performance of both methods aligns with the theoretical guarantees.

4 - Unraveling Temporal and Spatial Dynamics: An Interpretable St-Ode Method for Predicting Delivery TIME

Hao Wang, UC Berkeley, Berkeley, CA, United States, Junyu Cao, Yan Leng

We develop a new interpretable deep learning method by modeling continuous time on spatial networks to enhance the accuracy of delivery time prediction in home appliance retail sector. Driven by the operational complexities observed at RiRiShun Logistics, a prominent company in China’s home appliance delivery and installation industry, we introduce a novel model that integrates spatial network data within a continuous-time dynamic framework. This model is tailored to capture: (1) the irregular and sporadic patterns of dynamic order placements; (2) the time-dependent relationships within the spatial delivery network; and (3) the varying influences of different delivery centers. Through detailed numerical experiments on RiRiShun’s comprehensive dataset, we demonstrate that our model achieves superior accuracy in delivery time prediction compared to the current best-performing benchmarks. Our ablation study dissects the contribution of individual model elements, revealing how the variability in order flow, the dependencies within the delivery network, and the distinct roles of origin and destination centers can significantly influence predictive performance. The results indicate improvements up to 44.2% for different components, underscoring the efficacy of our motivation and model. These findings not only enhance the reliability of delivery predictions in the retail sector but also contribute a new perspective to the literature on dynamic prediction models, particularly highlighting the benefits of integrating continuous-time analysis with spatial network data.

MA54

Summit - 446

Bias in Online Reviews and Evaluations

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Hallie Cho, Vanderbilt University, Nashville

1 - The Differing Impact of Female Technical Leaders vs. Female Operational Leaders on Product Quality in the Medical Device Industry

George Ball, Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, IN, United States, Donggyu Jeon, Owen Wu, Gil Souza

Examining the influence of gender diversity in general, and female leadership in particular, on product quality performance in firms is a burgeoning research domain. While limited studies have examined if or how female leadership impacts product quality, no studies of which we are aware have differentiated between female technical leaders and female operational leaders. We do so in the highly technical medical device industry, in which product quality outcomes are primarily influenced by one of two underlying firm functions: design or manufacturing. Along those lines, we examine all publicly traded medical device firms in the U.S., and measure the proportion of female leadership team members that hold technical leadership positions, such as research, design and engineering vice-presidents, as well as the proportion of female leadership team members that hold operational leadership positions. These include roles such as manufacturing, supply chain and purchasing vice-presidents. We find that while in aggregate, adding more female leaders to a medical device firm reduces future serious product quality adverse events, this effect is completely explained by female technical leaders; female operational leaders appear to have no measurable impact on product quality complaints. Further, we employ detailed firm-level time-variant culture measures using Glassdoor.com, and discern that the impact of female technical leaders on quality is most prominent when a firm has a relatively negative overall culture, compared to a positive culture. Our study thus helps inform firms as to both when to add female leaders and in what roles to add them.

2 - A Moment For Reflection: De-biasing Service Evaluations

Hallie Cho, Vanderbilt University, NASHVILLE, TN, United States, Dawson Kaaua

This study examines different types biases in the evaluation of a service. Service evaluations typically come in two types--quantitative ratings and qualitative written feedback. Although ratings are typically collected before the written feedback, we propose that switching this order to ask for written feedback first can mitigate any bias present in star ratings as the process of writing provides time and space for reflecting on the evaluator's experience and analytical thinking. We conducted an experiment on Amazon Mechanical Turk where 4261 participants interacted with a server and evaluated their performance quality. We do not find strong effects of common biases toward servers (such as race and gender) in our setting but find strong effects of operational bias--sequence of good or bad experiences leads to biased star ratings. Most importantly, we find that collecting written feedback prior to star ratings effectively mitigates the detected sequence bias in subsequent star ratings. Our paper shows that most widely used quantitative feedback (i.e., average star ratings) can be biased if collected in the traditional manner and that this bias can be reduced if qualitative feedback is collected first. This work has major implications for any firm that collects quantitative and qualitative feedback and bases strategic decisions on aggregated quantitative feedback.

3 - The Impact of Political Polarization on Restaurant Reviews: Evidence from Yelp.Com

Rajan Mishra, California State University, Northridge, CA, United States, Sugato Chakravarty, Wreetabrata Kar

This study examines the influence of political polarization on online reviews and consumer behavior, using Yelp restaurant reviews from the 2016 election cycle. The bedrock of the Social Identity Theory is used to hypothesize that political shifts of a region from Democratic to Republican results in fewer, shorter reviews with simpler language and less praise. Our findings confirm these hypotheses, indicating that political polarization significantly impacts consumer behavior and online review language. Our research makes at least three important contributions. It empirically demonstrates the effect of political polarization on consumer behavior and language use; it underscores the potential impact of political polarization on the quantity and quality of online reviews; and it stresses the need for policymakers and online platform administrators to consider these effects and develop mitigation strategies. In as much as shorter reviews can convey less relevant/important information to the millions of consumers of online reviews, our study provides a timely wakeup call for all stakeholders.

4 - Decoding Biases: Unraveling Fairness and Discrimination in Human and Algorithmic Decision-Making

Arunima Chhikara, University of Kansas, Lawrence, KS, United States, Anurag Garg, Naveen Kumar, Liangfei Qiu

Decision-making exerts a profound influence on both individual destinies and broader societal trajectories. The advent of AI has been driven by the need for assistance in navigating intricate choices. While algorithmic systems promise heightened accuracy, they can inadvertently inherit biases from their underlying rules or training data, thereby engendering inequitable results. Discrimination becomes apparent when protected attributes, including race and gender, sway decisions, perpetuating a cycle of injustice. Despite the presence of numerous laws and policies designed to mitigate discrimination, both human and AI decision-making processes continue to grapple with biases. Our study delves into bias and discrimination in the context of online deception detection (ODD). We explore how gender and race bias affect the detection of fake reviews by AI systems and humans. Moreover, we conduct a randomized experiment to investigate taste-based and statistical discrimination in human decision-making in ODD and establish causal inference. Discerning various forms of discrimination can inform platform policies, bolstering trust. We draw on observational data from Yelp, complemented by randomized experiments, to offer a robust multi-method validation and glean valuable insights. Our research carries significant implications for platforms, policymakers, and businesses, advocating for fairness in online decision-making.

5 - Coping Through Precise Labeling of Emotions: A Deep Learning Approach to Studying Emotional Granularity in Consumer Reviews

Ali Faraji-Rad, University of Maryland, College Park, MD, United States, Ali Tamaddoni, Atefeh Jebeli

When describing their emotions, people may demonstrate *emotional expertise* by differentiating between emotions when using emotional labels or use emotion labels interchangeably to indicate a general valence. The authors develop a novel deep-learning-based method to measure the granularity with which people describe their emotions via language. They investigate the role of emotional granularity in consumer decision making, specifically in relation to coping with negative consumption experiences described in online reviews. Granularity in describing negative emotions is associated with more successful coping with negative experiences. Therefore, especially when the overall experience is negative, in which case coping is most relevant, greater granularity in describing *negative* emotions predicts more *positive* ratings of the business. Furthermore, in line with the view that the ability to granularly describe negative emotions is a skill, reviewers progressively become more granular when describing their negative emotions as they write more reviews. Consequently, reviewers progressively provide more positive ratings for negative experiences as they write more reviews. Finally, a greater temporal distance between the consumption experience and the writing of the review predicts greater granularity in describing negative emotions. Consequently, when the overall experience is negative and coping is relevant, a greater temporal distance predicts more positive ratings.

MA55

Summit - 447

Towards a Sustainable Resilient Future - Part A

Invited Session

Public Sector OR

Chair: Himadri Sen Gupta, University of Oklahoma, 202 W Boyd St, Norman, OK, 73019, United States

1 - Impact of Collaborative Food Rescue Initiatives on Food Waste Reduction

Ricky Owusu, North Carolina A&T State University, Greensboro, NC, United States, Lauren Davis

In the current food systems landscape, the concept of food rescue has become a powerful force due to concerns about waste and its effects on the environment and society. This paper explores the collaborative efforts of various stakeholders, such as retail donors, food banks, for-profit organizations, and waste management entities, in addressing the complex challenges of food security, sustainability, and waste reduction. Through a comprehensive framework, surplus food is rescued from various stages of the supply chain and redirected towards meaningful purposes, which helps mitigate the environmental impact of food waste. However, despite these efforts, the global paradox of food waste alongside increasing food insecurity continues to exist. This study examines the dynamics of how food is allocated among stakeholders through Monte Carlo simulation in order to identify disparities and optimize the use of resources. By addressing disparities and improving collaboration, the paper aims to enhance the efficiency and sustainability of food rescue initiatives, ultimately reducing food waste and improving societal well-being. Key research questions explore the impact of for-profit food rescue on non-profit organizations, the influence of optimized food allocation on sustainability, and the role of contracts in fostering collaborative efforts to combat food waste.

2 - A Disaster Management Digital Twin Framework for Resilient Electric Power Restoration

Abdullah Braik, Texas A&M University, College Station, TX, United States, Maria Koliou

Efficient and rapid restoration of electric power following natural hazards is crucial for community resilience, playing a pivotal role in emergency response, economic recovery, and the functionality of essential lifeline and social infrastructure systems. However, previous frameworks available in the literature have focused either on uncertainty modeling for pre-hazard mitigation and preparedness or on post-disaster data analysis for continuous learning and improvement. This study introduces a novel digital twin framework for immediate and rapid post-hurricane electric power restoration. The framework uses a hybrid approach that combines physics-based and data-driven models by utilizing a dynamic Bayesian network. By capturing the complexities of power system dynamics and incorporating the influence of the road network, the proposed framework provides a holistic approach to address restoration challenges. A discrete event simulation is conducted to showcase the framework's efficacy, demonstrating how the electric power restoration digital twin can be monitored and updated in real-time to reflect changing conditions, thereby facilitating risk-informed adaptive decision-making. Moreover, the framework's flexibility enables decision-makers to prioritize essential residential and business infrastructure and compare different restoration plans, assessing their potential impact on the community. In conclusion, the proposed framework holds the promise to revolutionize disaster recovery strategies, leading to faster recovery, improved emergency response, and enhanced community resilience.

3 - Exploring Interconnected Infrastructure Restoration and Resilience with Reinforcement Learning

Ribhu Sengupta, University of Washington, Seattle, WA, United States

With the Pacific Northwest falling along the Cascadia Subduction Zone fault running from Vancouver Island, Canada to Northern California, there is an increased probability that within the next 50 years, a disaster on the scale of a Magnitude 9 earthquake will occur. For a disaster of this magnitude, contingency plans for critical infrastructure networks will have to be developed for restoration efforts to help society recover. We model infrastructure restoration processes through Agent Based Simulation modeling at various levels of awareness/preparedness of interconnected infrastructure systems utilizing a sample setting with six subareas within a city sharing four infrastructure systems (Water, Ground Transportation, Electricity, and Telecommunications). To help guide the development of restoration plans for such networks, we explore the insights various Multi-Agent Reinforcement Learning (MARL) formulations can glean for infrastructure management officials in their planning efforts for the various levels of awareness. These MARL based approaches are benchmarked against theoretical optimal solutions for performance and validated using well known case study examples from the natural hazards resilience modeling platform IN-CORE. The findings from the MARL formulations are translated into actionable items for critical infrastructure managers to incorporate into their contingency planning.

4 - Mitigating Natural Hazard Impacts on Network Systems

Himadri Sen Gupta, University of Oklahoma, Norman, OK, United States, Andres Gonzalez, Charles Nicholson

Enhancing the resilience of utility networks against natural hazards is crucial for maintaining essential services during and after such events. This research presents a stochastic programming model to optimize the resilience of critical utility networks, specifically focusing on power grids and water supply systems. By incorporating the uncertainties associated with natural hazard occurrences and their impacts, the model generates probabilistic scenarios to assess potential disruptions. These scenarios are then used to optimize network configurations and response strategies, aiming to minimize service interruptions and expedite recovery times. The model is tested with data from Joplin, MO, a city severely impacted by natural hazards such as tornadoes. Preliminary results demonstrate significant improvements in network reliability and service continuity, validating the effectiveness of the proposed approach. This study contributes an optimization-based framework for enhancing the resilience of utility networks in hazard-prone areas.

MA56

Summit - 448

Life in Academia

Panel Session

Minority Issues Forum

Co-Chair: Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Moderator Panelist

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

2 - Panelist

Diana Ramirez-Rios, University at Buffalo, Buffalo, NY, United States

3 - Panelist

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

4 - Panelist

Arkajyoti Roy, The University of Texas at San Antonio, San Antonio, TX, United States

5 - Panelist

Chris Parker, Darden, Charlottesville, VA, United States

MA57

Summit - Terrace Suite 1

Healthcare Decision Models

Invited Session

Health Applications Society

Chair: Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States

1 - Supervised Inverse Optimization with Exogenous Decision Information

Felix Parker, Johns Hopkins University, Baltimore, MD, United States, Kimia Ghobadi

There are many potential practical settings of Inverse Optimization in which the typical strong assumptions on the forward problem structure and near-optimality of observations do not hold because the decision-making process was not optimized exactly, but exogenous information about the performance of observed decisions exists.

We propose Supervised Inverse Optimization (SIO), which aims to meld existing approaches for IO with supervised machine learning by incorporating scores for each observation to learn the objective function parameters through a model that optimizes both IO and regression objectives. The observation scores represent partial information about the objective function, which is often available, and reduces the reliance on assumptions, thereby improving performance. The model is highly flexible and can model problems with linear or non-linear objectives, objectives with unknown structure, multiple scores per observation, and many other problem settings.

SIO is valuable in real-world applications including diet recommendation, hospital capacity management, and finding shortest paths in traffic networks, among others. We demonstrate this with empirical results for the data-driven diet recommendation problem and for inferring and improving hospital capacity management strategies.

2 - Precision Staffing for Float Nurses Integrating AI and Optimization

Jonathan Helm, Indiana University, Bloomington, IN, United States

This talk describes an academia industry collaboration to design and execute a new internal travel nursing program at IU Health that moves flexible nurses between the 16 system hospitals to respond short term geographic fluctuations in demand. For adoptability, we developed decision support to create an on-call list of how many nurses will move and where 2 weeks in advance and deployment decisions 24-48 hours in advance. This required developing a machine-learning-based occupancy forecasting model that accounts for different levels of patient acuity and outputs distributional information that can be used to characterize stochastic nurse demand across the network that feed a stochastic optimization. Implemented in May 2023, subsequent 8 months indicates reductions of 17% understaffing, and 5% overstaffing. Annual savings >$750K.

3 - Neural Combinatorial Contextual Bandit Under Constraints for Sequential Resource Allocation

Ha Bui, Johns Hopkins University, BALTIMORE, MD, United States, Anqi Liu, Kimia Ghobadi

We study the Sequential Resource Allocation (SRA) problem, i.e., how to find an optimal decision of a fixed budget to minimize the allocation cost over time. We formalize this problem under the Combinatorial Multi-Armed Bandit (CMAB) setting. Then, we propose the Neural Combinatorial Upper Confidence Bound Resource Allocation (Neural-CUCB-RA) algorithm, a neural-network-based model to improve the allocation performance by learning from observed data over time. We show our method can work in the healthcare domain with a real-world COVID-19 dataset, in which the model allocates COVID-19 patients to different hospitals with different numbers of available ICU beds so that can save as many patients as possible.

4 - Enhancing Radiotherapy Selection through Predictive Modeling of Treatment-Induced Toxicities

Ibrahim Chamseddine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Radiation therapy is essential for over half of all cancer patients, requiring careful selection of treatment modalities based on patient-specific needs and available technologies. Photon therapy, the more common option, often exposes normal tissue to harmful radiation, making it less suitable for patients at high risk of side effects. On the other hand, proton therapy offers better protection for healthy tissues but is not as widely available. To optimize the allocation of proton therapy, we developed a deep learning model that predicts the risk of radiation-induced toxicities by incorporating patient-specific characteristics and radiation dose distributions in critical non-tumor tissues, referred to as organs-at-risk. The model, applied initially to liver cancer, was trained at Massachusetts General Hospital and validated with an external cohort from MD Anderson Cancer Center. This model has proven superior to traditional risk assessment methods, which typically rely on the average radiation dose delivered to the liver. By achieving high sensitivity in detecting high-risk patients and high specificity in confirming low-risk cases, the model facilitated the creation of personalized decision maps that aim to assist clinicians in selecting the most appropriate treatment modality. We further integrated the model directly into the treatment plan optimization process, through which we were able to adjust dose constraints in the liver based on the patient predicted risk, thereby personalizing treatment plans. Applied similarly to prostate cancer, our model demonstrated its ability to stratify patients into photon versus proton therapy based on their risk of rectal toxicity.

5 - A Network-Flow Approach to Increasing Inter-Facility Transfers for Nova Scotia Health Long-Term Care

Peter Vanberkel, Dalhousie University, Halifax, NS, Canada, Amelia Lane, Claver Diallo

Due to access and flow challenges within acute care, nearly 3,000 long-term care residents in Nova Scotia, Canada are on a waitlist to be transferred to their preferred facilities. Reasons for transferring include moving closer to families, cultural communities, and to reunite with a life partner in a different facility. We developed a new method to identify an optimal series of transfers which considers residents' care needs (i.e. required bed type, smoking, life partner), priority level, facility preferences, and the features of a vacant bed. We model this by generating a network of every resident on the waitlist and solving it as an elementary path problem, generating a chain of transfers starting with a vacancy. We use dynamic programming to solve this problem in practice. In this talk, the mathematical models and algorithms used to generate these chains, as well as an overview of their potential impact, will be discussed.

MA58

Summit - Terrace Suite 2

Junior Faculty Lightning Session

Panel Session

Health Applications Society

Co-Chair: Hrayer Aprahamian, Texas A&M University, College Station, TX, United States

1 - Moderator Panelist

Hrayer Aprahamian, Texas A&M University, College Station, TX, United States

2 - Panelist

Jackie Baek, NYU Stern School of Business, New York City, NY, United States

3 - Panelist

Karen Hicklin, Icahn School of Medicine at Mount Sinai, New York, NY, United States

4 - Panelist

Katherine Adams, University of Wisconsin-Madison, Madison, WI, United States, Justin Boutilier, Sarang Deo, Yonatan Mintz

5 - Panelist

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States

6 - Panelist

Karmel Shehadeh, Lehigh University, Bethlehem, PA, United States

7 - Panelist

Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

8 - Panelist

Zhaowei She, Singapore Management University, Singapore, Singapore

9 - Political Leadership and Contraceptive Uptake in India

Amir Karimi, The University of Texas at San Antonio, San Antonio, TX, United States

10 - Panelist

Shubranshu Shekar, Brandeis University, Waltham, MA, United States

11 - Panelist

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

12 - Panelist

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

MA60

Summit - Ballroom 2

Amazon Last-Mile Resource Optimization

Invited Session

The Practice Section of INFORMS

Chair: Liron Yedidsion, Amazon, Redmond, WA, United States

1 - Optimizing Tour Slicing via Shortest Path Formulations

Michael Wagner, University of Washington, Seattle, WA, United States, Dipal Gupta, Rohit Malshe

We present a novel method for slicing a Traveling Salesman tour into sub-tours that correspond to capacitated (both in time and space) vehicle routes. The slicing is guided by an optimization objective related to business metrics such as volume and time utilization as well as cost. Our key contribution is formulating the partitioning problem as a shortest-path problem on an induced network. When the metrics are non-negative, we demonstrate that the fast Dijkstra's algorithm can be used to solve the shortest path problem. Even when metrics can be negative, the slower Bellman-Ford algorithm can be applied. This shortest path approach is shown to be significantly faster than existing deterministic dynamic programming solutions. The methodology is general and can be applied to other sequence-slicing applications beyond vehicle routing.

2 - Logistics Operations with Combined Deliveries and Pickups

Rohit Malshe, Amazon.com, Kirkland, WA, United States, Abhilasha Katariya, Ram Thiruveedhi, Chinmoy Mohapatra

Combined deliveries and pickups are a logistics challenge. Highlighting the significance of the issue, we address the complexities faced in achieving optimal efficiency while ensuring customer satisfaction through shipper experience, promise, speed, and cost-effectiveness. Key sub-problem includes strategizing around the time lag between route planning and pickup windows to prevent over or under planning across multiple horizons under uncertainty.

3 - Newsvendor Model for Capacity Planning.

Liron Yedidsion, Amazon, Redmond, WA, United States, Rohit Malshe

The Newsvendor (NV) model calculates the target volume for each station that would minimize the cost expectancy under stochastic demand. The model considers the costs and capacities of each overage and underage scenario, In most cases, expected overage costs are higher than the expected underage costs, and thus the optimal target volume is usually smaller than P50. Using a segmented newsvendor logic, more volume is allocated to flexible means of delivery even though they are costlier per package.

4 - Essential components of capacity planning under uncertainity

Ram Thiruveedhi, Amazon, Leander, TX, United States, Abhilasha Katariya

Effective capacity planning of resources necessitates accurate estimates of demand and resource utilization: demand forecasting, resource usage forecasting, and capacity planning. We describe various processes to monitor and improve the effectiveness of the overall capacity planning process. These processes include tracking the accuracy of capacity forecasts, attributing errors to various forecast components, and continuous monitoring. Furthermore, we will explore the impact of underage and overage situations, where resources are either insufficient or excessive, respectively. These imbalances can lead to significant inefficiencies, such as lost revenue opportunities or unnecessary expenses. We will discuss solutions to address the imbalance and minimize the associated costs.

5 - Cycle Optimization Engine: Maximizing Last Mile Delivery Performance

Wen Chong, Amazon, Kirkland, WA, United States, Marc Anderson, Yimin Liu

Amazon orders are are fulfilled at a Fulfillment Center (FC) and shipped to a Delivery Station (DS) via line haul trucks. At the station, arriving packages are processed in a sequence of "cycles" -- operating shifts -- where packages are inducted and sorted into routes before dispatching to drivers for last mile delivery. Because packages bound for possibly faraway geographies arrive randomly throughout the day, with no cycle planning we potentially create inefficient routes with low delivery density and packages per route. Furthermore, some packages have high delivery sucesss rates at certain times of day (e.g. locker deliveries in the morning), which would be compromised if we neglect their time affinities. Finally, oversized (OV) packages are challenging to deliver, and we want to avoid repeatedly concentrating them in certain cycles/routes.
The novel Cycle Optimization Engine (COE) uses a MIP algorithm to compute the optimal cycle for each arriving package based on its delivery location, time-of-day preference and attributes such as OV, locker, business/residential. The algorithm maximizes each cycle's delivery density while respecting the station's complex processing constraints. It also prescribes the optimal use of buffering capacities for delaying packages to later, more favorable cycles. OV packages are dispersed both across different cycles and also geographically to avoid concentrating them on a small number of routes. In simulations, we show improvements in last mile delivery performance and financial savings.

MA61

Summit - Ballroom 3

How Do We Explain to OUr Neighbors What We Do

Panel Session

Committee’s Choice

Co-Chair: Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Tinglong Dai, Johns Hopkins University, Baltimore, MD, 21202, United States

Co-Chair: Sheldon Jacobson, University of Illinois, Urbana, IL, 61801, United States

1 - Moderator Panelist

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

2 - Panelist

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

3 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

MA62

Summit - Signature Room

Food Bank Operations: A U.S. Perspective on Humanitarian Food Assistance

Invited Session

TutORial

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Food Bank Operations: A U.S. Perspective on Humanitarian Food Assistance

Lauren Davis, North Carolina A&T State University, Greensboro, NC, United States, Irem Sengul Orgut, Steven Jiang, Eric Aft, Charlie Hale, Larry Morris, Jean Rykaczewski

Food banks are non-profit organizations with the primary mission of providing food

assistance to the communities they serve. This assistance occurs through the collec[1]tion, storage, and distribution of food to people in need through a complex supply

chain network of donors and charitable agencies. Distribution of food is challenging

in this environment, given the many resource constraints experienced at the finan[1]cial, human, and material levels. Furthermore, the COVID-19 pandemic highlighted

the important role these organizations play as demand for food assistance surged,

while sources of supply were constrained. This tutorial provides an overview of food

bank operations from a supply chain perspective. We specifically characterize the key

stakeholders, product, and information flows within the food bank supply chain and

draw from our prior experience with several U.S. food banks to delineate structural

differences that exist among these supply chain networks. We further elucidate the

influence of the supply chain network design on the organizations' operational deci[1]sions and strategic direction with respect to equity, efficiency, effectiveness, diversity

and inclusion. We present several studies that illustrate the role of descriptive, pre[1]dictive, and prescriptive analytics in improving the distribution of food and reducing

food waste, as well as provide insights for future research in this area.

MA63

Regency - 601

Emerging Topics in IS

Invited Session

Information Systems

Chair: Mingwen Yang, University of Washington, Seattle, WA, United States

1 - Helping Oneself by Helping a Rival: Effects of a Social Media Platform & Rsquo;S Decision to Enable Features that Help Influencers

Yan Zhu, University of Texas at Dallas, Richardson, TX, United States, Jianqing Chen, Srinivasan Raghunathan

Social-media platforms and influeners have long been seen as collaborators. However, the fact that both of them make money through advertising services adds a competitive dimension to their relationship. While some platforms have adopted an antagonistic approach, other have adopted a friendly approach in allowing the influencers to use features that are available to the platform. We develop a game-theoretic model to study the platform's decision to enable a feature for influencers' use. The influencer's service becomes more effective and more efficient if she uses the feature. We show that the platform would be better off enabling the feature if the advertisers' valuation of the platform's service is not too low compared to that of the influencer, if the feature's valuation-enhancement effect is not too low nor too high, and the cost-reduction effect is not high. Under those conditions, while the platform loses market share when it enables the feature, the loss is offset by the higher price the platform charges. Our results imply that whether the platform should provide help hinges critically on the nature and extent of the help. Additionally, while giving the influencer access to the feature could be a win-win situation for both the platform and the influencer, their gains are at the expense of some advertisers, suggesting the platform's helping the influencer can lead to a collusive outcome and can act as a advertiser surplus extraction strategy, despite the price competition between them for advertisers.

2 - Competing with Generative AI: Does It Hurt Creative Content Producers?

Zhen Yan, University of Texas at Dallas, Richardson, TX, United States, Jianqing Chen, Srinivasan Raghunathan

This study investigates the impacts of Generative Artificial Intelligence (GenAI) on the creative industry, focusing on the competition between the expert and nonexpert creators. Based on a game-theoretic model, we examine how GenAI reshapes creators' strategic allocation of budget with respect to the routine technical-execution dimension and the nonroutine originality dimension. GenAI has two main impacts: automating the routine tasks and augmentation creators in doing nonroutine tasks. We find whether the expert could benefit from GenAI depends critically on whether the expert's capability advantage over the non-expert is greater on the originality dimension or the execution dimension. If the capability advantage is greater on the originality dimension, the expert is always better off with GenAI. On the other hand, if the capability advantage is greater on the execution dimension, the expert benefits from GenAI only when consumers' appreciation of originality vis-à-vis execution is not significant, and GenAI's augmentation effect dominates the automation effect.

3 - Flexible Estimation of Network Models

Ye Liu, University of Washington, Seattle, WA, United States

The curse of dimensionality issue is at the center of analysis with network data, where traditional empirical models either resort to aggregated measures at the node level or require extraordinary computational effort as the size of the network grows. Recent developments in machine learning have introduced numerous neural network structures essential for network analysis, such as Graph Convolutional Networks (GCN). However, these neural networks typically treat network links and node characteristics as separate inputs, failing to capture the interplay between node characteristics and network structure. In this paper, we propose a scalable, yet simple neural network structure as a non-parametric method to estimate models with network data as input. We demonstrate through extensive simulations that our proposed functionals can flexibly capture the underlying interplay between the network links and node characteristics in a completely data-driven fashion. We also apply our method to a real-world dataset to demonstrate the practicality of our method, showing its ability to handle large-scale networks efficiently.

4 - Win-Curves in Real-TIME Bidding: a Micro-Foundation

Shailender Joseph, University of Texas at Dallas, Richardson, TX, United States, Ganesh Janakiraman, Milind Dawande, Vijay Mookerjee

The emergence of digital media has transformed the advertising landscape. Real-time bidding (RTB) has become an important avenue for selling the digital ad space that is generated in these media outlets. The RTB process is a set-up where the publishers can auction off each impression via the supply-side platform (SSP), and the demand-side platform (DSP) can make bids for these impressions so that they can complete the ad campaigns on behalf of the advertisers. The interactions between the SSP and DSP are facilitated by an ad exchange.

A win-curve in this context is a function that tells the amount to bid in order to achieve the win (i.e., the probability of winning for a given bid) probability that the bidder desires. A sizable number of previously published works that studied the RTB process assumed the existence of such a curve to derive their results. Our work provides a formal and rigorous framework on how such a win-curve emerges in an RTB setup. This work is a step in this direction where we attempt to understand the microstructure of RTB that helps in the characterization of the equilibrium in this setup, and the emergence of win-curve is a part of this characterization. We also provide a method to compute this win-curve and implement this method on some numerical examples to understand its shape.

MA64

Regency - 602

Social Media Analytics on E-Commerce

Invited Session

Social Media Analytics

Chair: Jae Hong Park, Kyung Hee University, Dongdaemun-gu

1 - Impact of Financial APIs on the Trade Outcome of Retail Investors

Bongjin Sohn, Korea University, Seoul, Korea, Republic of, Ziru Li, Gunwoong Lee

This research investigates the impact of financial Application Programming Interfaces (APIs) on retail investors' trading behavior in the stock market. Traditionally, accessing financial information necessitated extensive manual procedures. The advent of financial APIs, however, has remarkably simplified the retrieval and analysis of corporate data for investors. Initially exclusive to institutional investors, financial APIs now grant retail investors access to the same data. This study aims to examine the impact of financial APIs on retail investors, leveraging the introduction of the OpenDART service in South Korea. OpenDART provides company disclosure information via APIs. Analyzing trading volume, net-buying trends, and stock returns of retail investors around OpenDART' launch, we find an increase in retail investors' trading volume and a decrease in the trade quality post API introduction. For lesser-known stocks, moreover, we observe a more pronounced increase in abnormal trading volume and a notable decline in trade quality. The findings contribute to the literature on electronic disclosure systems, financial technologies, and the behavior of retail investors, offering managerial implications for both retail investors and policymakers in the stock market.

2 - Performance after Reward, Penalty, and Break-Even: Empirical Study with mHealth App under Reward-Penalty Incentive Structure

Kyung Pyo Kang, Kyung Hee University, Seoul, Korea, Republic of, Seungwook Jin, Keumseok Kang, Jae Hong Park

The reward-penalty structure is a composite incentive system used in mobile health platforms, requiring users to deposit money before starting goal-oriented activities. At the end of the activities, users can receive rewards, face penalties, or get refunds depending on their performance. Previous research suggests that rewards and penalties can positively motivate users by encouraging them to seek rewards or avoid penalties. However, their effects may extend beyond a current activity, influencing performance in subsequent activities either positively or negatively. For example, individuals might either improve performance after facing penalties to recover losses or perform poorly due to decreased motivation from negative experiences. Understanding the effects of rewards and penalties on users' subsequent activities is essential for platform managers, because the primary goal of mobile health platforms is to maintain user motivation throughout their goal-oriented activities. Nevertheless, previous studies have mainly focused on the effects of incentive structures on single activity, overlooking the long-term incentive effects on performances with respect to the continuous stream of goal-oriented activities. While previous research on financial incentives has explored the sole effects of either rewards or penalties, scant literature examines how the reward-penalty composite structure works. This study addresses the literature gaps by examining the dynamics of goal achievement within the reward-penalty incentive structure, specifically in the context of a mobile health (mHealth) application. We investigate how rewards, deposit refunds, and penalties influence users' subsequent short-term and long-term goal achievement rates. We expect this paper will enhance understanding of the mechanisms underlying reward-penalty incentive structure.

3 - The Substitution Effect of YouTube Videos on Netflix Viewership: An Empirical Study on the Impact of Video Length and Popularity

Seungwook Jin, KAIST College of Business, Seoul, Korea, Republic of, Keumseok Kang

In the era of digital content, user-generated content (UGC) in the form of video has become an important source of information for consumers. While previous literature has primarily focused on the promotional aspect of UGC, this study extends the conventional view by examining the substitution effect of video UGC on over-the-top (OTT) service. Drawing on Uses and Gratifications Theory (UGT), we hypothesize that the gratifications derived from video UGC can substitute for the utility of consuming original content on OTT platforms, ultimately cannibalizing its viewership. We investigate the impact of video UGC on the viewership of Netflix's top 10 movies by analyzing a dataset of 315 movies and their associated weekly view hours and ranks, as well as related video UGC posted on YouTube. We find that while the overall effect of weekly UGC volume on movie viewership is positive, suggesting a promotional effect, this relationship is negatively moderated by video length, likeability, and number of views. These findings suggest that as video UGC becomes longer, more liked, and more viewed, it provides consumers with a higher level of satisfaction, leading to a substitution effect on viewership of the original film on the OTT platform. Our findings have important implications for content creators and OTT platform managers, highlighting the need to strategically manage video UGC to optimize its promotional benefits while mitigating its potential substitution effects on viewership.

MA65

Regency - 603

Power of LLMs in the Analytics Classroom: The Good, the Bad, and the Inevitable I

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Sourav Chatterjee, University of North Texas, 1155 Union Circle, #311160, Denton, TX, 76203, United States

Co-Chair: Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, United States

1 - Beyond Search: How LLMs are Transforming Informal Learning and Code Development

Erin Trochim, University of Alaska Fairbanks, Fairbanks, AK, United States

Large Language Models (LLMs), the technology powering conversational AI tools like ChatGPT and Gemini, are changing how we learn and create. While early experiences with chat boards and search engines often required iterative strategies to find applicable information, modern LLMs have unlocked a new era of possibilities. As LLMs are reshaping the landscape of informal learning, we will focus on implications for creating materials for and learning coding and analytics.

We'll explore how LLMs can enhance code standards by generating clear examples, optimizing existing code, and providing insightful explanations. This approach offers the opportunity to refine code accessibility and enhance understanding. We will compare models with specific custom instructions and examples versus using a prompt engineering chat approach. We'll also examine how these tools can support self-guided learning, offering on-demand explanations, tutorials, and problem-solving assistance.

2 - How ChatGPT can support learning from a student's perspective

Zoe Long, University Of Florida, Gainesville, FL, United States, Elif Akcali

ChatGPT is one of, if not, the most used natural language processing tools in many industries. With ChatGPT's increasing popularity, it is important to integrate it into the educational system to prepare students for the workforce. Students need to be able to understand how and when to use large language processing models to be successful in their careers. An assignment was created for students in an upper-level industrial and systems engineering class during the spring 2024 semester. For the assignment, students were asked to prompt ChatGPT to create an essay with multiple references about ABC analysis and then to analyze each sentence of the essay for correctness of the technical information and the relevance of references cited. At the end of the semester, students were asked to reflect on their experience with the assignment and how it impacted their learning. Of the 38 students in the class 31 consented to have their ChatGPT generated essays, their analysis of their text, and their responses to reflection prompts used for research purposes. These student responses to the reflection prompts

were qualitatively analyzed using open and axial coding. This presentation will focus on some preliminary insights on how students feel using ChatGPT can be used to support their educational experience, and how it impacts their understanding of technical concepts.

3 - Beyond Spreadsheets: Embracing ChatGPT in Business Analytics Education

Mustafa Hayri Tongarlak, Ivey Business School, London, ON, Canada, Mehmet Begen, John Wilson

Analytics courses on decision-making are popular in business schools. Most such courses use spreadsheets and add-ins as their modelling tools. Recent developments in large language models (LLM), such as ChatGPT, can allow instructors to use LLM agents instead of spreadsheet-based tools to enhance student interaction and learning and focus more on Decision-Making with Analytics (DMA) skills. We critically analyze the implications of this adoption in terms of benefits and challenges, in addition to ethical considerations, potential pitfalls, and practical recommendations. Through case studies and feedback from pilot implementations, we aim to showcase the tangible benefits and encourage a broader discussion on optimizing AI integration in DMA education.

4 - The challenges and opportunities of integrating LLMs in classrooms

Hayley Falk, University of Michigan, Ann Arbor, MI, United States

Large language models (LLMs), powered by artificial intelligence (AI), have existed for decades, but were not widely known outside of academic circles. However, the launch of ChatGPT (Chat Generative Pre-Trained) in November 2022 by Open AI catapulted the field from the research community to the general public. Since November 2022, a significant number of LLMs have been developed with seemingly endless capabilities. Here we present a comparative analysis of teaching practices employed in teaching a Python programming curriculum to high school students pre- and post-ChatGPT. Our analysis will highlight some potential challenges and opportunities of incorporating AI language models in classrooms and in reshaping educational paradigms.

MA66

Regency - 604

AI in Mining and Mineral Processing

Invited Session

Artificial Intelligence

Chair: Bart Maciszewski, Imperial Oil, Calgary, AB, Canada

1 - Offline Reinforcement Learning for Chemical Process Control with a Digital Twin

Dhavalkumar Patel, University of Toronto, Toronto, ON, Canada, Stefano Scaini , David Molina Concha, Chi-Guhn Lee

The application of Reinforcement Learning (RL) for process control has attracted much attention in process-driven industries. RL is an area of machine learning that is focused on learning optimal decisions through experience. Given that RL is traditionally an online paradigm, that is, data-based experiences are gained by interacting with real processes in real time, its adoption for uses that are prohibitively dangerous or costly has been limited. The chemical processing industry in particular is an industry defined by highly controlled and risk-intensive extraction and manufacturing processes which do not lend themselves naturally to the use of RL, despite the great potential for unlocking value that RL offers to many chemical processing applications. The field of Offline Reinforcement Learning (Offline RL) has emerged largely as an answer to this difficulty, with the goal of learning being redirected to fixed batches of recorded data. This eliminates the need to interact with live processes directly, and when combined with supplemental models of the process environment—digital twins—this approach can lead to strong outcomes. The talk will focus on the use of Offline RL in combination with process simulation modelling to develop an optimal control policy for pH stability.

2 - Enhance Mine Operations Through Analytics

Hannah Zhao, Imperial Oil Limited, Calgary, AB, Canada, Rahul Raman, Saied Afshari, Andrew Milne, Dmitry Yakunin, Farah ElMallah, Anuraj Grewal, Jessica Pickel, Richard Leger, Freeman Lin

In the dynamic landscape of mining operations, leveraging data analytics and machine learning has become paramount for productivity and operational reliability gains and informed decision-making. At Imperial's Kearl oil sands mine, we've harnessed these technologies to drive meaningful improvements.

By analyzing shovel sensor data, we've gained insights into effective shovel operation. Our custom dashboard breaks down shovel operators' performance, aiding targeted training, ultimately helping increase productivity, and lower operating and maintenance costs.

We've tackled shovel crowd cylinder failure analysis using anomaly detection techniques. Hydraulic crowd cylinder failure can be costly and causes unplanned downtimes, so understanding when failure happens is essential. Our approach, utilizing autoencoders, provides valuable insights for maintenance strategies and component longevity.

Furthermore, we currently have a manual process to infer tracked vehicle undercarriage components replacement timing. We are developing models to aid in predictive maintenance by analyzing historical data, usage patterns, and wear rates.

3 - Applications of Artificial Intelligence for Advancing Short-Term Mine Planning Leveraging Real-TIME Sensing Technologies

Zachary Levinson, AIRL Technologies, North Vancouver, BC, Canada, Julien Keutchayan

At large-scale mining operations worldwide, sensors play a crucial role by providing rich data on orebody characteristics and equipment performance. Leveraging this data, an optimization framework is developed to enhance short-term production plans by learning to adapt to real-time operating conditions over time. This framework is tested at a copper mine, where it demonstrates significant improvements in productivity and more efficient extraction of valuable materials. These results underscore the potential of integrating sensing data into production planning to boost overall operational efficiency and resource utilization.

4 - Reinforcement learning for optimizing underground mine planning and design considering geotechnical, geological, and grade uncertainty

Jeff Boisvert, University of Alberta, Edmonton, AB, Canada, Roberto Noriega, Yashar Pourrahimian

Uncertain geological and mineral grade estimates due to scarce drill-hole data present a key challenge in the design and strategic planning of underground mining projects. Geological complexities and variations in mineral grades undermine the reliability of mine plans and designs, often leading to suboptimal production strategies, heightened financial risks and missed opportunities. Numerous geostatistical techniques have been developed to assess uncertainty in mineral estimates and provide multiple simulated scenarios based on the current level of knowledge of the deposit, however, incorporating these simulated models within a mine planning and design framework remains a major challenge. In this work, the application of Reinforcement Learning (RL), an Artificial Intelligence (AI) technique that deals with the problem of decision-making in uncertain environments, is explored to incorporate an ensemble of simulated mineral deposit scenarios in the generation of underground reserves and mine extraction plans. A risk-discounted strategy is presented that penalizes the financial risk of the mine design and allows a user specified attitude towards risk, a trade-off between expected profit and financial risk over the ensemble of geological and grade scenarios. A case study based on the extraction of a gold deposit using sublevel stoping method is presented to highlight the impact of geological and grade uncertainty on the financial performance of an underground mining asset. The proposed methodology incorporates an ensemble of mineral resource realizations in the design and extraction planning, respecting operational and geotechnical constraints, to aid practitioners make informed choices in the face of uncertainty.

5 - Fault-tolerant Autonomous Control System

Mahmut Tatlici, University of Alberta, Edmonton, AB, Canada, Oguzhan Dogru, Biao Huang

Primary Separation Cells (PSCs) are crucial in bitumen extraction, separating oil sands slurry into bitumen froth, middlings, and coarse tailings. Optimizing the froth-middling interface is vital for maximizing recovery and minimizing environmental impact. However, designing an autonomous control system for PSCs presents challenges like operating-point optimization, interface tracking, and fault tolerance. To address these challenges, we propose a machine learning-based fault-tolerant control system using three reinforcement learning (RL) agents to enhance a model predictive controller (MPC)-based control loop. The first agent uses camera images for accurate interface tracking, the second tunes the MPC for optimal set-point tracking, and the third optimizes the set-point for maximum recovery and minimal environmental impact. The MPC controller, supervised by these agents, calculates optimal control actions and completes the feedback loop. Despite these improvements, faults such as actuator issues, sensor failures, and communication errors can necessitate human intervention. To address this, we implement a sensor fusion approach with a Kalman filter-based prediction model to correct interface measurements during faults. Experimental validation on a pilot-scale PSC demonstrates the approach's effectiveness. Our results show the potential of machine learning, combined with fault-tolerant techniques, to achieve autonomous control in complex systems. This research advances control algorithms in nonlinear systems, providing a general solution for optimizing bitumen extraction and other similar environments requiring fault-tolerant autonomous control.

MA67

Regency - 605

AI in Financial Ethics and Fraud Detection

Contributed Session

Contributed

Chair: Xingying Wu, Xi'an JiaoTong university, 西安市, N/A

1 - Perspectives on Artificial Intelligence Integration in Accounting

Shari Fowler, Indiana University East, Richmond, IN, United States, Marcy Jance

The presentation discusses a survey that will be sent to accountants to gauge accountants’ views on how artificial intelligence is impacting current and future accounting jobs and how artificial intelligence can be used to prevent fraud, financial reporting misstatements, tax evasion, and other ethical violations.

2 - Using Artificial Intelligence and Analytics Techniques to Combat Accounting Fraud

Marcy Jance, Indiana University East, Richmond, IN, United States, Shari Fowler

The presentation discusses the types of accounting fraud that may be uncovered by using artificial intelligence techniques. The presentation then discusses how artificial intelligence and analytics can be used to help companies detect accounting fraud in financial statements.

3 - Firm-level AI Ethics Awareness Measurement and Effects: A Machine-Learning Approach

Xingying Wu, Xi'an JiaoTong university, 西安市, China, People's Republic of, Shan Liu

Recent years have witnessed the rapid development of artificial intelligence (AI) and its pervasive impacts on business and society. However, the deployment of AI also brings ethical challenges regarding beneficence, privacy, and security. A thorough understanding and effective management of AI ethics issues are imperative to ensure the socially beneficial development of AI. Drawing on the AI ethics frameworks outlined by the National Overall Group for Artificial Intelligence Standardization, we employ textual analysis techniques to develop a novel firm-level, time-varying measure of AI ethics awareness (AIEA), which captures the attention management devotes to AI ethics related issues. We validate AIEA through content analysis, variance decomposition, and by considering the time series and cross-industry variations, which ensures that the measure is specifically indicative of the intended construct. Subsequently, we identify its determinants and reveal that firms with larger size, greater Environmental, Social and Governance (ESG) commitment, and managers with higher educational levels, can improve firms' AIEA level. We investigate how AIEA affect firms' innovation behavior, revealing that firms with higher AIEA have higher digital innovation. We also discover that firms with higher AIEA enjoy lower audit pricing and positive market reactions, suggesting that the disclosure of AIEA serves as a signal of firms' ethical management and mitigates the perceived business risk of the auditor and investors. Our research contributes to the AI ethics literature by developing a novel AIEA measure with broad applicability and offers practical implications for firms to enhance their awareness and voluntary disclosure on AI ethics issues.

MA68

Regency - 606

Hydrogen Production and Emission Strategies

Contributed Session

Chair: Rui Zhou, University of Tennessee, Knoxville, knoxville, TN, United States

1 - DRI Operation, Costs, and CO_2 Emissions under Different Energy Markets

Simon Kamerrer, TU Dortmund, Dortmund, Germany, Elina Hoffmann, Valerie Karplus, Chris Pistorius, Christian Rehtanz

In response to the climate crisis, steelmakers have increasingly turned to direct reduction (DR) to achieve their CO_2 emissions reduction targets. This study explores the operational flexibility and economic viability of three DR technologies under different economic conditions to investigate how variation in local energy markets may drive technology choice. Hydrogen direct reduction (H-DR) eliminates fossil fuel use but requires robust hydrogen supply chains or onsite hydrogen electrolysis, whereas natural gas direct reduction (NG-DR) leverages existing natural gas infrastructure and can be coupled with carbon capture and sequestration (CCS) for deeper decarbonization. Hybrid-DR allows flexibility between hydrogen and natural gas fuels based on economic factors and availability. Our time-series based optimization incorporates detailed material balances to simulate the costs and emissions of these different DR processes, economically optimizing the operation of each plant layout and quantifying the costs associated with CO_2 avoidance. We find that the NG-DR with CCS and Hybrid-DR plants offer cost advantages through flexible operation, minimizing the use of their decarbonization technologies, whereas the H-DR plant achieves consistently low CO_2 emissions but is highly sensitive to electricity prices. We find that the NG-DR with CCS and H-DR plants are able to achieve lower mean costs at higher emissions reduction targets, with H-DR dominating in regions of low electricity prices and high natural gas prices. This research underscores the importance of examining market environments, operational strategies and technological flexibility to inform investment and policy decisions for decarbonizing the steel industry.

2 - Advancing the Technology Adoption for an Effective Scale Up of Green Hydrogen

Jingwei Chen, University of Tennessee, Knoxville, Knoxville, TN, United States, Mingzhou Jin

Green hydrogen represents a promising frontier in sustainable energy, offering a zero-emission alternative to fossil fuels crucial for decarbonization efforts worldwide. Despite its potential, scaling up green hydrogen production faces numerous challenges, including high costs of electrolyzer technologies, substantial renewable energy requirements, integration complexities within existing energy infrastructures, and regulatory hurdles. Market dynamics and technological uncertainties further complicate the adoption process. To address these issues effectively, we introduce a novel approach using stochastic programming to develop an optimization model tailored to manage uncertainties. This model aims to navigate cost variability, fluctuating energy supplies, and evolving policy landscapes, providing practical solutions to enhance the capacity and scalability of green hydrogen production, thus accelerating its integration into the global energy mix.

3 - Hydrogen Logistics Network Design for Hydrogen Fuel Cell Electric Long-haul Trucks

Rui Zhou, University of Tennessee, Knoxville, knoxville, TN, United States

With the growing emphasis on reducing carbon emissions and transitioning towards sustainable transportation solutions, hydrogen fuel cell electric long-haul trucks have emerged as a promising alternative to conventional diesel-powered vehicles. This study addresses the critical need for designing effective hydrogen delivery networks for long-haul trucks under different demand. Through the utilization of advanced modeling techniques and optimization algorithms, we seek to identify optimal configurations for hydrogen delivery technologies (compressed gas, liquid hydrogen, and ammonia) and refueling infrastructure. The outcomes of this research are expected to provide valuable insights into the strategic planning of hydrogen logistics networks, thereby facilitating the widespread adoption of hydrogen fuel cell electric long-haul trucks and contributing to the decarbonization of the transportation sector.

MA69

Regency - 607

Digital Technologies in Business

Invited Session

eBusiness

Chair: Kai Sun, University of Texas at Dallas, Richardson, TX, United States

1 - How Do Human Mental Health Care Providers React to Their AI Counterparts? a Study of An Online Counseling Forum

Gujie Li, University of Maryland, College Park, MD, United States, Jui Ramaprasad, Lauren Rhue

Mental health care has become a concern globally, and the challenges of providing and accessing appropriate care is only magnified by a dearth of available counselors. At the same time, Large Language Models (LLMs) are anticipated to hold transformative potential for numerous industries including healthcare. LLM's exceptional capacity to offer individually-tailored recommendations derived from massive medical knowledge can facilitate access to affordable mental health services. Naturally, an inquiry arises regarding how human counselors respond to the emergence of AI counselors. We focus on an online counseling platform in China wherein professional counselors and peers offer support to mental health care seekers through question-and-answer (Q&A). In February 2023, an LLM-powered chatbot was integrated into the platform to address inquiries from seekers, providing an excellent context for us to study changes in behavior of incumbent human counselors in the presence of an AI counselor. Specifically, leveraging the (unanticipated) introduction of the chatbot as a natural experiment, we investigate how human counselors react to the AI counselor in terms of the level of engagement, quality of assistance and counseling strategies.

2 - Visibility OR Capability: An Economic Analysis of Livestreaming Adoption

Xiang Ji, University of Science and Technology of China, Hefei, China, People's Republic of, Qingning Cao, Min Chen

Livestreaming marketing is increasingly becoming a popular strategy to promote, market, and sell products in online retailing. This study uses a game-theoretic framework to provide insights into the firm's livestreaming strategy.

3 - An Economic Analysis of Regulation on Clinical Algorithmic Fairness

Zhe Wang, City University of Hong Kong, Hong Kong, Hong Kong, Dengpan Liu

In the context of healthcare, employing a game-theoretical model, we examine the impact of algorithmic fairness regulation on key market outcomes, including patient surplus, healthcare providers' payoffs, and social welfare. Our analytical results show that, when healthcare algorithm fairness regulations become stricter, the performance of algorithms among majority patient groups worsens, leading to a decrease in utility for majority patient groups. Meanwhile, surprisingly, when fairness regulations for algorithms are overly stringent, further reinforcing fairness restrictions may even worsen the performance of algorithms among minority patient groups and potentially reduce the utility of minority patient groups. Additionally, our research findings indicate that as fairness regulations become stricter, healthcare service providers initially increase and then decrease their healthcare service fees. Considering the impact of fairness regulations on diagnostic accuracy for different groups and healthcare service fees, we find that appropriately emphasizing algorithmic fairness can improve overall diagnostic accuracy and potentially enhance overall patient utility and societal welfare, however, excessively emphasizing algorithmic fairness can reduce overall diagnostic accuracy and harm overall patient utility and societal welfare. Based on this, our analysis further addresses how to regulate healthcare algorithm fairness policies. Specifically, our research results suggest that for policymakers aiming to maximize social welfare, the optimal algorithm fairness constraints should relax as effort cost parameters increase or co-insurance rates increase, but tighten as the baseline diagnostic accuracy of minority patient groups increases or the severity of symptoms increases.

4 - Using Optimization to Augment Structural Estimation: a Reinforcement Learning Approach

Kai Sun, The University of Texas at Dallas, Richardson, TX, United States, Wangsheng Zhu, Vijay Mookerjee

Structural Models have been used in many business domains including Information Systems, Finance, Marketing, and Operations Management. One advantage of structural models is that they can be used to perform counterfactuals or policy simulations that predict what would happen if certain elements of the economic environment were to change. However, the validity of these predictions critically depends on the assumptions made to build the model. We could have competing models that fit the observed data but make different predictions. Re-enforcement learning (RL) allows us to find near-optimal solutions to complex problems without making structural assumptions. We demonstrate using this property of RL to choose between different models that statistically fit the data.

MA70

Regency - 701

Power Systems on the Road Between Optimization and Learning

Invited Session

ENRE: Electricity

Chair: Vladimir Dvorkin, University of Michigan, Ann Arbor, 48104

1 - Value-Oriented Renewable Energy Forecasting

Yuanyuan Shi, University of California San Diego, La Jolla, CA, United States, Yufan Zhang, Honglin Wen, Yuexin Bian

Energy forecasting is an essential task in power system operations. Operators usually issue forecasts and leverage them to schedule energy dispatch ahead of time. However, forecast models are typically developed in a way that overlooks the operational value of the forecasts. To bridge the gap, we design a value-oriented forecasting approach for sequential energy dispatch problems with renewable energy sources. Specifically, the forecast model parameter estimation is formulated as a bilevel program, optimizing the forecast model parameter at the upper level and the operation decisions are determined at the lower level. The DA and RT market clearing problems are solved at the lower level, with the RES forecast from the upper level serving as the input. Theoretically, we derive the exact form of the loss function for training the forecast model that aligns with such a goal. For market clearing modeled by linear programs, this loss function is a piece-wise linear function. Additionally, we derive the analytical gradient of the loss function with respect to the forecast, which inspires an efficient training strategy. Numerical study shows our forecasts can bring benefits of the overall cost reduction compared to quality-oriented forecasting.

2 - Dynamic Optimal Power Flow: Real-TIME Stochastic Predictive Control Through Constrained Reinforcement Learning

Anna Scaglione, Cornell University, New York, NY, United States, Tong Wu

Traditional Deep Reinforcement Learning (DRL) techniques have shown promise in addressing control challenges within power systems that are associated with the volatility and stochasticity introduced by high penetration of renewable resources and responsive demand. However, vanilla DRL does not ensure compliance with crucial physical constraints like power flow equations and voltage limits. Existing approaches typically project actions onto feasible sets, potentially leading to sub-optimal solutions. In this presentation we introduce a novel primal-dual method to train constrained DRL policies tailored for real-time stochastic dynamic optimal power flow control. We also prove the convergence of critic and actor networks and showcase the effectiveness of our approach through case studies on IEEE standard systems, highlighting its ability to identify near-optimal actions while maintaining safety compliance.

3 - Learning Context-Aware Security Prescriptions in Optimal Power Flow

Robert Mieth, Rutgers University, Piscataway, NJ, United States

The massive roll-out of stochastic resources (e.g., wind and solar power) in the power system complicates its operations and challenges established regulatory and economic paradigms. Yet, while novel and provably effective decision-making techniques can internalize parameter stochasticity efficiently, they are often complex and deviate too much from existing approaches for a timely implementation in practice. This talk discusses approaches on how optimal power flow, a central tool for power system operations and planning, can be enriched with few interpretable parameters that are prescribed such that its performance is as close as possible to a fully stochastic approach. We briefly survey the state-of-the-art and present a novel data-driven approach to learn an efficient parameter prescription for robust optimal power flow.

4 - Diffusion Based Electric Vehicle Charging Modeling

Yize Chen, University of Alberta, Edmonton, AB, Canada

Recent proliferation of electric vehicle (EV) charging load has imposed vital stress on power grid. The stochasticity and volatility of EV charging behaviors render it challenging to manipulate the uncertain charging demand for grid operations and charging management. Charging scenario generation can serve for future EV integration by modeling charging load uncertainties and simulating various realistic charging sessions. To this end, we propose a denoising Diffusion-based Charging scenario generation model coined DiffCharge, which is capable of yielding both battery-level and station-level EV charging time-series data with distinct temporal properties. In principle, the devised model can progressively convert the simply known Gaussian noise to genuine charging demand profiles by learning a parameterized reversal of the forward diffusion process. Besides, we leverage the multi-head self-attention mechanism and prior conditions to capture the unique temporal correlations associated with battery or charging station types in actual charging dynamics. Moreover, we validate the superior generative capacity of DiffCharge on a real-world dataset involving ample charging session records, and attest the efficacy of produced charging scenarios on a practical EV operation problem in the day-ahead electricity market.

5 - Regression Nash Equilibrium in Electricity Markets

Vladimir Dvorkin, University of Michigan , Ann Arbor, MI, United States

We argue that the economic efficiency of renewable-dominated electricity markets improves if stochastic producers coordinate on their forecast models. To support this argument, we introduce the ***regression equilibrium***—a state of the electricity market where no profit-seeking stochastic producer benefits by unilaterally deviating from their respective forecast model. While the equilibrium aims to maximize the private welfare, i.e., the profit of each market participant, it also proves to align with the socially optimal solution that minimizes the average electricity cost across the day-ahead and real-time stages, thus improving temporal market coordination. We base the equilibrium analysis on the theory of variational inequalities, providing results on the existence and uniqueness of such an equilibrium in energy-only markets, and deriving two methods for equilibrium computation.

MA71

Regency - 702

Developing Machine Learning Methods for Complex Data Distributions and Structures

Invited Session

Data Mining

Chair: Yanjun Qian, Virginia Commonwealth University, Glen Allen, VA, United States

1 - Machine Learning-Based Risk Adjusted Cusum Control Chart for Monitoring Surgery Readmission Rate

Yanjun Qian, Virginia Commonwealth University, Glen Allen, VA, United States

Percutaneous Transhepatic Biliary Drainage (PTBD) Catheter Placement is known to suffer from a high readmission rate. This work focuses on monitoring the change in readmission rate after a new clinical paradigm for post-procedural care is implemented for the PTBD procedures. The Risk-Adjusted CUmulative SUM (RA CUSUM) chart, considering the preoperative risk of patients, is widely used for similar tasks. However, the popular logistic regression model for RA CUSUM charts in the literature shows a weak discriminant ability for the readmission prediction of the PTBD procedures. To overcome these limitations, we incorporate RA CUSUM charts with other machine learning methods for modeling the patients' risks based on their characteristics. We use data retrospectively and prospectively collected before and after the paradigm change from the medical college of Virginia Commonwealth University to evaluate the performance of the proposed method. The RA CUSUM charts based on tree-based models show better discriminant ability and faster detection than the logistic regression one.

2 - Optimal Supervised Data Splitting

Samuel Paul Gyamfi, Virginia Commonwealth University, Richmond, VA, United States, Sujay Mudalgi, Anh Bui

Data splitting is an important step in predictive modeling. State-of-the-art data splitting methods accomplish this via minimizing the energy distance between the split and original data. However, they do not specifically consider the relationship between the response and predictor variables. In this article, we propose Alyke, a supervised data splitting method that aims to minimize the energy distance while considering this relationship. We demonstrate that Alyke can help improve predictive power in simulation and real examples. Alyke is also much faster than the state-of-the-art data splitting methods, making it a more viable solution for big data applications. We also discuss the application of Alyke in supervised data compression problems.

3 - Predicting Dengue Incidence in the Dominican Republic Using Climate Data

Sahil Chindal, Virginia Commonwealth University, Richmond, VA, United States, Demian Herrera, Manuel Colome Hidalgo, Helena Sofia Rodrigues, Yanjun Qian, Michael Robert

Dengue is a mosquito-borne disease prominent in tropical and subtropical regions of the world and has been emerging in temperate areas. Dengue is endemic to the Dominican Republic, where outbreaks have been occurring for the last four decades. Most provinces in the

Dominican Republic have a tropical climate with abundant rainfall during rainy seasons, providing ample resources for mosquito proliferation and growth, which creates an environment favorable for dengue transmission. Using climate data and dengue case data, we aim to determine which climate factors are associated with dengue cases. Additionally, due to the potential lags between climate and dengue trends, we seek to characterize the lags between variables associated with climate and cases. Using machine learning methods, we analyze the temporal dynamics of dengue spread by estimating the parameters of the SIR-type model framework. We validate our model using the historical data for dengue cases in the Dominican Republic. We will use our results as part of a predictive model to forecast the spread of dengue cases between various provinces of the Dominican Republic and integrate our model into an early warning system that predicts outbreaks and informs public health and mosquito control policies in the Dominican Republic.

4 - Out-of-Distribution Learning with Gaussian Process Boundaries

Wenbo Sun, University of Michigan Transportation Research Institute, Ann Arbor, MI, United States, Yang Chen, Chih-Li Sung, Arpan Kusari

Deep neural networks (DNNs) are usually constructed under the closed-world assumption, which is violated in real-world tasks, due to presence of out-of-distribution (OOD) samples. This leads to DNNs outputting highly overconfident wrong predictions and can lead to disastrous consequences in safety-critical applications. All existing OOD detection methods rely on curating a set of OOD samples to train the models or perform hyperparameter tuning to separate the in-distribution (InD) samples from the OOD, failing to achieve a high detection accuracy without exposure to OOD samples in the training phase. Given that humans can reason about what they do not know through a process called Metacognition, we propose a Gaussian-process-based OOD detection method to establish a detection boundary based on InD samples only. The basic idea is to perform uncertainty quantification of the unconstrained softmax scores via a clustered Gaussian process, and then define a score function to separate InD and OOD samples based on their fundamental differences in the posterior predictive distribution from the Gaussian process. The method is shown to have good statistical properties. The comparison study results demonstrate that the proposed method consistently outperforms the state-of-the-art OOD detection methods.

5 - Riemannian Manifold Approach to Time Varying Mixture Models of Object Shapes

Rifat Hasan, Florida State University, Tallahassee, FL, United States, Chiwoo Park

Geometrical forms of products are important quality characteristics in advanced manufacturing, e.g., geometries of 3D printed products and morphology of nano-manufactured products. Hence, the importance of quantifying the distribution of the geometric forms of objects along with evaluating the quality and variation of related manufacturing processes. Here, we present a new time varying mixture model for the probability distribution of geometrical forms in 2D. Based on Riemannian geometry, an appropriate space of shapes and proper metric over the space is defined. After that, we present the time varying mixture model with Gaussian-type distributions. Here, we consider that only mixture proportions will change along with time; the other parameters like mean and covariance will be constant. Finally, the numerical result of the algorithm is provided with applications to modeling the time varying mixture models of nano-products.

MA72

Regency - 703

Machine Learning for Heterogeneous Data Fusion

Invited Session

Data Mining

Chair: Hyunsoo Yoon, Yonsei University, Seoul, Korea, Republic of

1 - Balanced Cross-Domain Fault Diagnosis Method in Rotating Machinery with Imbalanced Data

Yongmin Kim, Yonsei University, Seoul, Korea, Republic of, Hyunsoo Yoon

Traditional unsupervised domain adaptation-based bearing fault diagnosis research assumes that the source machine and target machine data have the same label distribution. However, in real-world manufacturing process, most of the data from the target machine is normal, with few or no labeled fault classes, and even unlabeled fault classes often have only a few samples. This class imbalance leads to poor performance since unsupervised domain adaptation models tend to overfit to the abundant normal data and do not effectively learn from the limited defect data. Thus, class imbalance makes it challenging for typical unsupervised domain adaptation methods to generalize for diagnosing faults on the target machine. To address this issue, we propose an unsupervised domain adaptation technique to tackle the problem of class imbalance in target machine data. Our approach aims to prevent overfitting to normal data and enhance the model's learning capability from insufficient defect data. As a result, the proposed method improves fault diagnosis performance, ensuring more reliable and accurate identification of bearing faults in real-world manufacturing scenarios.

2 - Fine-Tuning Run-to-Run Process Control Method Using An Autoencoder for Chemical Mechanical Polishing

HYEONGGU LIM, Yonsei University, Seoul, Korea, Republic of, Chang Ouk Kim

In the chemical mechanical polishing (CMP) process of semiconductor manufacturing, run-to-run (R2R) process control is critical for consistently producing high-quality products. R2R process control involves optimizing the process parameters of equipment to maintain high-quality products despite process noise such as equipment aging. In this paper, we propose a novel method to overcome two limitations of existing R2R process control algorithms when applied to CMP processes. The first limitation is that ratios between control variables are not considered, despite being essential. The second is that it is not possible to prioritize between control variables when there are multiple ways to reach a target. The proposed method can update the control variables while maintaining the ratios between the control variables, enabling continuous process control even in severe noise. The cost of process control can be reduced by setting priorities among the control variables to reach the target. In addition, the proposed method estimates the changes in the process environment with an autoencoder, achieving better performance on several process control metrics than other R2R algorithms.

3 - Leveraging Deep Domain Adaptation for Improved Estimation of Remaining Useful Life

Youngje Oh, Yonsei University, Seoul, Korea, Republic of, Hyunsoo Yoon

Prognostics and Health Management (PHM) is extremely important in industrial environments. In particular, the prediction of Remaining Useful Life (RUL) has gained significant attention due to its impact on equipment operating costs and safety. While most studies employ supervised learning models for training and evaluation, relying solely on training data from specific domains can lead to performance limitations. The variability and imbalance in data across different domains can degrade the generalization capability of prediction models. This study aims to introduce a methodology to enhance the performance of RUL prediction models by enriching the source domain through data augmentation and utilizing deep domain adaptation to create a model more suited to the target domain. Additionally, by applying methods tailored to the characteristics of sensors to extract features, we aim to improve prediction performance and reduce variability. This methodology enhances the practical applicability in industrial settings and is expected to significantly improve the efficiency and safety of equipment operations.

4 - Multiclass Classificaiton of Alzheimer's Disease Prodromal Stages Using Sequential Embeddings and Regularized Multikernel Support Vector Machine

Oyekanmi Olatunde, Amazon, Seattle, WA, United States, Kehinde Oyetunde, Jihun Han, Mohammad Khasawneh, Hyunsoo Yoon

The detection of patients in the cognitive normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) stages of neurodegeneration is crucial for early treatment interventions. However, the heterogeneity of MCI data samples poses a challenge for one-shot CN vs. MCI vs. AD multiclass classification, as some samples are closer to AD while others are closer to CN in the feature space. Previous attempts to address this challenge produced inadequately accurate results, leading most frameworks to break the assessment into binary classification tasks such as AD vs. CN, AD vs. MCI, and CN vs. MCI. Other methods proposed sequential binary classifications such as CN vs. others, and then dividing others into AD vs. MCI. While those approaches may have yielded encouraging results, the sequential binary classification method makes interpretation and comparison with other frameworks challenging and subjective. Those frameworks exhibited varying accuracy scores for different binary tasks, making it unclear how to compare the model performance with those of other direct multiclass methods. Therefore, we introduce a classification framework comprising unsupervised ensemble manifold regularized sparse low-rank approximation and regularized multikernel support vector machine (SVM). This framework first extracts a joint feature embedding from MRI and PET features, which were then combined with the Apoe4, Adas11, MPACC digits, and Intracranial volume features using a regularized multikernel SVM. The framework achieved a state-of-the-art (SOTA) result in a CN vs. MCI vs. AD multiclass classification and generalizes well to binary classification tasks, achieving SOTA results in all but the CN vs. MCI category.

5 - A Multi-Source Domain Generalization Method for Intelligent Fault Diagnosis Based on Domain Meta-Information

Hoyeon Lee, Yonsei University, Seoul, Korea, Republic of, Hyunsoo Yoon

Intelligent fault diagnosis plays a crucial role in ensuring the reliability and efficiency of complex machines in industrial fields. However, traditional fault diagnosis methods often fail to generalize well to new or unseen machines due to domain discrepancies among machines operating under different conditions. Domain adaptation aims to mitigate such discrepancies by adapting models trained on a source domain to perform effectively on a target domain. While effective, domain adaptation typically needs access to data from the target domain during training, which may not always be feasible or cost-effective. In contrast, domain generalization approaches focus on learning representations that are invariant to domain discrepancies, without requiring any data from the target domain. The proposed method extends the capabilities of domain generalization by leveraging the meta-information from multiple source domains. We evaluate the effectiveness of the proposed method on various benchmark fault diagnosis datasets. Experimental results demonstrate that our method outperforms other existing domain generalization methods, achieving superior fault diagnosis accuracy.

MA73

Regency - 704

INFORMS Journal on Data Science (IJDS) Spotlight Session

Invited Session

Data Mining

Chair: Nick Street, University of Iowa, Iowa City, IA, United States

1 - Causal Decision Making and Causal Effect Estimation Are Not the Same…and Why It Matters

Foster Provost, New York University, New York, NY, United States, Carlos Fernández-Loría

Causal decision making (CDM) at scale has become routine. Increasingly CDM is based on machine learning. Businesses algorithmically target offers, incentives, and recommendations to affect consumer behavior. Recently, research related to CDM and causal effect estimation (CEE) using ML models has accelerated. However, CDM is not the same as CEE, and counterintuitively, accurate CEE is not necessary for accurate CDM. Our experience is that this is not well understood by practitioners or most researchers. Technically, the estimand of interest is different, and this has important implications both for modeling and for the use of statistical models for CDM. This paper highlights three implications. (1) We should consider carefully the objective function of the causal ML, and if possible, optimize for accurate treatment assignment rather than for accurate effect-size estimation. (2) Confounding affects CDM and CEE differently. For supporting CDM it may be just as good or even better to learn with confounded data. Finally, (3) causal modeling may not be necessary at all to support CDM because a proxy target for statistical modeling might do as well or better. This third observation helps to explain at least one common CDM practice that seems wrong at first blush---the widespread use of non-causal models for targeting interventions. The last two implications are particularly important in practice, as acquiring (unconfounded) data on both ``sides" of the counterfactual can be quite costly and often impracticable. To facilitate research in this area we point to recent related articles from multiple contributing fields.

2 - GIFAIR-FL: A Framework for Group and Individual Fairness in Federated Learning

Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States, Xubo Yue, Maher Nouiehed

In this paper we propose GIFAIR-FL: a framework that imposes Group and Individual FAIRness to Federated Learning settings. By adding a regularization term, our algorithm penalizes the spread in the loss of client groups to drive the optimizer to fair solutions. Our framework GIFAIR-FL can accommodate both global and personalized settings. Theoretically, we show convergence in non-convex and strongly convex settings. Our convergence guarantees hold for both i.i.d. and non-i.i.d. data. To demonstrate the empirical performance of our algorithm, we apply our method to image classification and text prediction tasks. Compared to existing algorithms, our method shows improved fairness results while retaining superior or similar prediction accuracy.

3 - Achieving Reliable Causal Inference with Data-Mined Variables: A Random Forest Approach to the Measurement Error Problem

Mochen Yang, University of Minnesota, Minneapolis, MN, United States

Combining machine learning with econometric analysis is becoming increasingly prevalent in both research and practice. A common empirical strategy uses predictive modeling techniques to "mine" variables of interest from available data, then includes those variables into an econometric framework to estimate causal effects. However, because the predictions from machine learning models are inevitably imperfect, econometric analyses based on the predicted variables likely suffer from bias due to measurement error. We propose a novel approach to mitigate these biases, leveraging the random forest technique. We propose employing random forest not just for prediction, but also for generating instrumental variables for bias correction. The random forest algorithm performs best when comprised of a set of trees that are individually accurate in their predictions, yet which also make "different" mistakes, i.e., have weakly correlated prediction errors. A key observation is that these properties are closely related to the relevance and exclusion requirements of valid instrumental variables. We design a data-driven procedure to select tuples of individual trees from a random forest, in which one tree serves as the endogenous covariate and the others serve as its instruments. Simulation experiments demonstrate its efficacy in mitigating estimation biases, and its superior performance over alternative methods.

4 - Sparse Density Trees and Lists: An Interpretable Alternative to High-Dimensional Histograms

Lesia Semenova, Duke University, Durham, NC, United States, Siong Thye Goh, Cynthia Rudin

Histograms, as piecewise constant density estimation models, are easy to visualize, are accurate with enough data in each bin, and have a logical structure that most people find interpretable. However, all of these properties fail in more than two or three dimensions, particularly for binned binary or categorical data. We present sparse tree-based and list-based density estimation methods as higher dimensional analogies to variable bin width histograms for binary/categorical data. In each leaf of the tree (or list), the density is constant, similar to the flat density within the bin of a histogram. Our models can easily be visualized in more than two dimensions, are sparse, and have priors that help with generalization. Our generative modeling methods allow the user to specify the preferred number of leaves or branches within a Bayesian prior for the tree or the preferred number and length of rules within the prior for the rule list. The new approaches often yield a better balance between sparsity and accuracy of density estimates than other methods for this task. We present an application to crime analysis, where we estimate how unusual each type of modus operandi is for a house break-in.

MA74

Regency - 705

Power System Operation and Planning Under Uncertainty

Invited Session

ENRE: Electricity

Chair: Angela Flores, Universidad de Chile, Santiago, Chile

Co-Chair: Rodrigo Moreno, University of Chile, Santiago, Chile

1 - Power system planning under multi-scale climate-induced uncertainty

Angela Flores, University of Chile, Santiago, Chile, Iván Pérez, Byron Castro

Power systems are facing increasing pressure from climate change. The rise in global temperatures, more variable precipitation patterns, and the increase in frequency and severity extreme of weather events pose substantial risks for power systems in multiple time-scales. In this study we introduce a comprehensive power system planning model that allows to anticipate both short and long-term uncertainties, leveraging multi-scale stochastic programming. Furthermore, we present a novel approach for constructing multi-scale scenario trees for representing climate-induced uncertainty, using historical weather data and future climate projections.

2 - Analyzing and modeling the behavior of data centers and digital currency miners in power systems

Ignacio Andrés Aravena Solís, Lawrence Livermore National Laboratory, Livermore, CA, United States, Subir Majumder, Le Xie

Over recent years, power grids have seen an increase in the deployment and interconnection of large data centers and digital currency mining facilities, with power consumption levels reaching 700MW. In this work, we study the empirical behavior of large data centers and mining facilities in Texas and how their consumption is affected by factors like currency conversion factors, electricity price, local temperature, system demand, and different pricing mechanisms (seasons with and without peak pricing). We employ data transformation techniques to account for the very skew distribution of consumption (and demand response), and we fit and validate a SARIMA model to the transformed data. We find that during short-term mining operations, surprisingly, miners' consumption is not correlated or explainable by changes in digital currency prices. Instead, the primary factors impacting consumption are temperature, energy prices, and demand response strategies implemented in the Electric Reliability Council of Texas (ERCOT). We use the fitted model to generate public, synthetic datasets, which can be used to account for the behavior of this type of consumer in other markets. Finally, we discuss potential explanations for the observed behaviors and postulate profit-maximizing models that capture these behaviors for short-term operation and planning studies.

3 - Unit Commitment Problem with Energy Storage Under Correlated Renewables Uncertainty

Rodrigo Moreno, University of Chile, Santiago, Chile, Felipe Cordera, Fernando Ordonez

The extensive integration of renewable generation in electricity systems is significantly increasing the variability and correlation in power availability and the need for energy storage capacity. This increased uncertainty and storage capacity should be considered in operational decisions such as the short-term unit commitment (UC) problem. In this work, we formulate a day-ahead UC problem with energy storage, considering multistage correlated uncertainty on renewables' power availability. We solve this multistage stochastic unit commitment (MSUC) problem with integer variables in the first stage using a new variant of SDDP that can explicitly deal with temporal correlations. Our computational results on the IEEE 118-bus system demonstrate the significance of considering multistage uncertainty and correlations, comparing our solution with other multistage solutions, two-stage solutions, and deterministic solutions typically used by industry. We also solve the MSUC problem for a representation of the Chilean power system, finding superior UC solutions for scenarios where adapting generation to the unfolding uncertainty is costly. Finally, we demonstrate that the MSUC approach can be used to define a more efficient deterministic UC solution, outperforming the current industry practice.

4 - Capacity Expansion Planning under Uncertainty subject to Expected Energy Not Served Constraints

Marilena Zambara, National Technical University of Athens, Athens, Greece, Daniel Avila, Anthony Papavasiliou

We present a method for solving a large-scale stochastic capacity expansion problem which explicitly considers reliability constraints, in particular constraints on expected energy not served. Our method tackles this problem by a Lagrange relaxation of the expected energy not served constraints. We solve the relaxed formulation in an iterative manner, using a subgradient-based method. Each iteration requires the solution of a stochastic capacity expansion problem, for which we implement a subgradient decomposition scheme in a high-performance computing infrastructure. We apply the proposed methodology on the Economic Viability Assessment model that is used by ENTSO-E in the annual European Resource Adequacy Assessment. The Economic Viability Assessment is extended to include reliability constraints. We are able to solve this generalization of the Economic Viability Assessment and obtain the least-cost capacity expansion plan for meeting specific reliability requirements per zone, fully accounting for uncertainty. Our approach also allows us to compute the shadow price of load shedding by zone (through the Lagrangian multipliers of the relaxed constraints) which can serve as an indication of the zone-specific VOLL or price cap that could in principle deliver the appropriate level of investments within an energy-only market.

MA75

Regency - 706

Advancements in Learning and Control for Power Systems

Invited Session

ENRE: Energy-Climate

Chair: Apurv Shukla, Texas A&M, College Station, TX, United States

1 - Recent Advances in design of dispatch policies for power grid

Apurv Shukla, Texas A&M, College Station, TX, United States

Look-ahead dispatch is a widely deployed policy across system operators for economic dispatch. However, little analysis of the impact of inter-temporal ramping constraints, forecast errors, and window size on loss in terms of generation cost and price fluctuations have been studied in antecedent literature. In this paper, we provide counter-examples demonstrating system behavior contrary to established folklore practices. Building on these examples, we conduct a thorough performance analysis and develop performance guarantees for look-ahead dispatch policies in the presence of intermittent renewable generation, storage, network congestion, and load forecast errors. These insights are further used to provide several policy recommendations regarding the practical usage of such policies. All our results are corroborated on a suite of numerical experiments based on IEEE 24 and 118 bus-systems.

2 - Towards Closing the Electricity-Transportation-Human Loop via Control and Market

Junjie Qin, Purdue University, West Lafayette, IN, United States

Electric power systems and transportation systems become deeply interconnected due to transportation electrification. The couplings between these systems can be leveraged to optimize the performance of both systems. In this talk, we will go through two such examples. In the first example, we consider the design of workplace electric vehicle (EV) charging tariffs in a morning commute setting. We demonstrate that a time-varying charging tariff can play the role of a dynamic toll to improve the traffic or nudge the EV arrival times to optimize electricity system costs. In the second example, we introduce a business model where EVs are shared for joint provision of transportation and electricity services. We justify using shared EVs to provide a particular electricity service (i.e., demand charge reduction) by exploring the statistics of load data of large electricity users and analyze the optimal spatial pricing problem for joint service provision with a shared EV fleet. Impacts of the proposed operation is simulated using real transportation network company data and electric load data for the city of San Francisco.

3 - The Architecture of Green Energy Systems

Michael Ferris, University of Wisconsin, Madison, WI, United States, Andy Philpott

The world is undergoing a transition from using fossil-fuel energy that emits greenhouse gases (mainly carbon dioxide) to using energy that does not. This transition is a global response to calls to limit global warming that has been caused by the emission of greenhouse gases over the post-industrial era. The current scale and speed of this transition appears insufficient to keep global temperatures below agreed targets. There are many technical, economic, social and political reasons for this slowness that have been canvased in a number of recent reports.

Our purpose in this paper is to examine the contribution that mathematics and mathematical models can make to understanding and overcoming the barriers that are faced in the transition. In particular we will focus on what we call the architecture of energy systems, which consists not only of the physical infrastructure for generating and transporting energy, but also the market and contractual arrangements that give incentives for investing in this infrastructure and that allow for it to be operated in an efficient manner. This paper examines the contributions that mathematical modeling can make to help accelerate this transition. The models we catalog are confined to optimization and equilibrium models, but cover a range of physical scales and time scales. Our focus is on novel model formulations that can help overcome the challenges of the transition by unpicking the complexity inherent in many settings and quantifying the tradeoffs that must be made when developing energy policy.

4 - Optimal Forest Management for Interdependent Products: a Nested Stochastic Dynamic Bioeconomic Model

Tong Wu, Cornell University, Ithaca, NY, United States, C.-Y. Cynthia Lin Lawell, David Just, Jiancheng Zhao, Zhangjun Fei, Qiang Wei

Sustainable forest management is an important issue worldwide. Forests supply the world’s population with timber and non-timber forest products, including renewable products such as fruits, nuts, and maple syrup that can be harvested at more frequent intervals than the trees themselves. We develop a nested dynamic bioeconomic model and dynamic structural econometric model of the management of forests that generate interdependent products that differ in their growth cycles, rates of growth, lengths of growing periods, and potential harvest frequency. We apply our model to detailed daily panel data we have collected and constructed on bamboo shoot and bamboo stem harvesting decisions made by bamboo farmers, in order to assess the optimality of bamboo farmers’ forest management strategies and to understand the beliefs and perceptions that underlie and rationalize their management strategies. Our nested stochastic dynamic model nests an inner finite-horizon within-year dynamic programming problem that captures daily bamboo shoot growth within a season, inside an outer finite-horizon between-year dynamic programming problem that captures bamboo stem growth from year to year. Sources of uncertainty in our model include stochastic rain, stochastic prices, and stochastic survival. We plan to use our dynamic model to simulate, analyze, and design policies and institutions to improve sustainable forest management. Our research has important implications for the sustainable management of forests worldwide, particularly when the forests produce products that grow on trees, are renewable, and can be harvested at more frequent intervals than the trees themselves.

5 - Modeling Equilibrium EV Charging Behavior Considering Driver (Un)availability and Waiting Times

Hannah Davalos, UC Berkeley, Berkeley, CA, United States, Shmuel Oren, Candace Yano

Range anxiety and shortages of chargers in certain geographic areas are concerns for prospective EV purchasers and thus also for auto manufacturers. Using data from a U.S. car manufacturer, we examine actual charging patterns of EV drivers and develop a model that allows us to capture factors such as driver (un)availability, electricity time-of-use rates and disutility for waiting to obtain estimates of equilibrium EV charging behavior and waiting times. Sensitivity analysis using our model enables us to explore how the addition of more EV drivers and charging stations would affect the equilibrium day-of-week and time-of-day demands for charging and the load on the electrical grid.

MA76

Regency - 707

Energy Systems Modeling and Optimization VI: Methods for Modern Power Systems

Invited Session

ENRE: Other Energy

Chair: Antoine Lesage-Landry, Polytechnique Montréal, Montreal, QC, H4A 3L3, Canada

1 - Frequency Dynamics-Aware Real-TIME Locational Marginal Pricing of Electricity

Christine Chen, University of British Columbia, Vancouver, BC, Canada, Bo Chen, Roohallah Khatami, Abdullah Al-Digs

We present a method to generalize the well-known concept of locational marginal prices (LMPs) to embed the impact of system frequency dynamics into real-time electricity markets. The proposed frequency dynamics-aware LMPs can help to mitigate costs associated with setting aside ever more reserve capacity to offset larger, faster, and more frequent transient excursions arising from greater renewable integration. We formulate a dynamics-aware economic dispatch (ED) by augmenting a traditional static ED with constraints pertinent to system frequency dynamics, including those from inertial response, primary frequency control, and the automatic generation control. Similar to their traditional counterparts, dynamics-aware LMPs are composed of Lagrange multipliers associated with the power balance and transmission line power flow constraints. Numerical simulations involving standard test systems validate our findings, confirm added revenue opportunities for generators contributing to frequency support, and demonstrate computational scalability.

2 - Optimal Electric Vehicle Charging with Dynamic Pricing, Customer Preferences and Power Peak Reduction

Miguel Anjos, University of Edinburgh, Edinburgh, United Kingdom, Luce Brotcorne, Gaël Guillot

We consider a provider of electric vehicle charging that operates a network of charging stations and uses time-varying pricing to maximize profit and reduce the impact on the electric grid. We propose a bilevel model with a single leader and multiple disjoint followers. The customers (followers) makes decisions independently from each other. The provider (leader) sets the prices for each station at each time slot, and ensures there is enough energy to charge. The charging choice of each customer is represented by a combination of a preference list of (station, time) pairs and a reserve price. The proposed model takes thus into accounts for the heterogeneity of customers with respect to price sensitivity and charging preferences. We define a single level reformulation based on the reformulation for the rank pricing problem. Computational results highlight the efficiency of the new reformulation and the impact of the model on the grid peak.

3 - Online Interior-Point Methods for TIME-Varying Equality-Constrained Optimization

Jean-Luc Lupien, UC Berkeley, Berkeley, CA, United States, Iman Shames, Antoine Lesage-Landry

An important challenge in the online convex optimization (OCO) setting is to incorporate generalized inequalities and time-varying constraints. The inclusion of constraints in OCO widens the applicability of such algorithms to dynamic but safety-critical settings such as the online optimal power flow (OPF) problem. In this work, we propose the first projection-free OCO algorithm admitting time-varying linear constraints and convex generalized inequalities: the online interior-point method for time-varying equality constraints (OIPM-TEC). We derive simultaneous sublinear dynamic regret and constraint violation bounds for OIPM-TEC under standard assumptions. For applications where a given tolerance around optima is accepted, we propose a new OCO performance metric - the epsilon-regret - and a more computationally efficient algorithm, the epsilon OIPM-TEC, that possesses sublinear bounds under this metric. Finally, we showcase the performance of these two algorithms on an online OPF problem and compare them to another OCO algorithm from the literature.

4 - The Influence of Short-Term Electricity Pricing Strategies on Long-Term Supply Flexibility

Han Shu, Cornell University, Ithaca, NY, United States, Jacob Mays

As the rapid integration of renewable energy sources such as wind, solar, and storage, electricity markets demand to adopt pricing strategies to ensure long-term flexibility and grid reliability. This study investigates how short-term electricity pricing strategies influence long-term investments in flexible generators and energy storage. We develops a sequential decision problem faced by system operators, to access price formation in wholesale electricity markets given variability, uncertainty, non-convexity, and intertemporal operating constraints. Numerical study shows that different pricing strategies can result in markedly different price outcomes and revenues streams. Moreover, these pricing mechanisms shape the energy mix and grid reliability over time by either promoting or hindering the deployment of flexible generation resources. Our findings highlight the critical need for policymakers to integrate considerations of long-term flexibility into the design of short-term pricing policies, particular in the context of decarbonized electricity markets.

MA77

Regency - 708

Interfaces Between Simulation and Optimization

Invited Session

Computing Society

Chair: Shuotao Diao, Northwestern University, Evanston, IL, 60208, United States

1 - Multi-stage Stochastic Programming for Integrated Network Optimization in Hurricane Relief Logistics and Evacuation Planning

Sudhan Bhattarai, Clemson University, Clemson, SC, United States, Yongjia Song

This study addresses the integrated hurricane relief logistics and evacuation planning (IHRLEP) problem, combining hurricane evacuation and relief item pre-positioning operations, which are usually managed separately. We introduce a fully adaptive multi-stage stochastic programming (MSSP) model and develop solution approaches based on two-stage stochastic programming (2SSP). By leveraging historical forecast errors using an auto-regressive model of order one, we generate hurricane scenarios and approximate the hurricane process as a Markov chain, with each state defined by the hurricane's location and intensity. Our comprehensive numerical experiments, inspired by Hurricane Florence and Hurricane Ian case studies, demonstrate the superior out-of-sample performance of the MSSP model's fully adaptive policies compared to the static policies of the 2SSP model. Additionally, an extensive sensitivity analysis provides valuable insights into the conditions under which fully adaptive policies outperform static ones, highlighting key problem parameters.

2 - Convergence and Complexity Analysis of a Stochastic Frank - Wolfe Algorithm with Away Steps

Natthawut Boonsiriphatthanajaroen, Cornell University, Ithaca, NY, United States, Shane Henderson

We study the convergence properties of the Frank Wolfe algorithm, also known as the conditional gradient method, with away steps for linearly constrained stochastic optimization. The algorithm makes use of unbiased gradient estimates. The convergence analysis of the algorithm relies on an assumption that the estimated gradient is sufficiently accurate with high enough probability, which can be ensured in our context by using a sufficient number of replications. Under these and other conditions, we show that the algorithm converges to a neighborhood of a globally optimal solution where the rate of convergence, sample size, and number of iterations until convergence also depend on the geometry of the feasible set.

3 - Piecewise Linear Decision Rules are Asymptotically All You Need

Yue Lin, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Daniel Zhuoyu Long, Viet Anh Nguyen, Jin Qi

Two-stage risk-averse distributionally robust optimization (DRO) problems are ubiquitous across many engineering and business applications. In these problems, the decision-maker commits to a capacitated first-stage decision, anticipating that second-stage recourse decisions can be executed after observing the realization of the uncertain parameters. Despite their promising resilience, two-stage DRO problems are often intractable because the recourse decisions are infinite-dimensional. To address this issue, a practical yet powerful solution approach is to approximate the recourse decisions using parametric decision rules. This paper focuses on the approximation using piecewise linear decision rules (PLDRs), which deliver better quality than linear decision rules but are more computationally demanding. Besides distributional uncertainty, we also capture the decision maker's attitude towards risk. We propose a heuristic method to find the first-stage decision and the second-stage PLDR by solving a deterministic, finite-dimensional optimization problem. Moreover, our simple heuristic solution is asymptotically optimal as the magnitudes of the problem parameters, including the uncertain parameters and the first-stage capacity, inflate to infinity at an appropriate rate. This asymptotic optimality highlights the surprising power of PLDRs from a new perspective. We validate our theoretical findings through numerical simulations in assemble-to-order problems. In the data-driven setting, our finely tuned heuristic solution can achieve small optimality gaps and surpass the sample-average approximation method in the out-of-sample performance.

4 - Deterministic and Stochastic Frank-Wolfe Recursion on Probability Spaces

Raghu Pasupathy, Purdue University, West Lafayette, IN, United States

Motivated by applications in emergency response and experimental design, we consider smooth stochastic optimization problems over probability measures supported on compact subsets of the Euclidean space. With the \emph{influence function} as the variational object, we construct a deterministic Frank-Wolfe (dFW) recursion for probability spaces, made especially possible by a lemma that identifies a ``closed-form'' solution to the infinite-dimensional Frank-Wolfe sub-problem. Each iterate in dFW is expressed as a convex combination of the incumbent iterate and a Dirac measure concentrating on the minimum of the influence function at the incumbent iterate. To address common application contexts that have access only to Monte Carlo observations of the objective and influence function, we construct a stochastic Frank-Wolfe (sFW) variation that generates a random sequence of probability measures constructed using minima of increasingly accurate estimates of the influence function. We demonstrate that sFW's optimality gap sequence exhibits $O(k^{-1})$ iteration complexity almost surely and in expectation for smooth convex objectives, and $O(k^{-1/2})$ (in Frank-Wolfe gap) for smooth non-convex objectives. Furthermore, we show that an easy-to-implement fixed-step, fixed-sample version of (sFW) exhibits exponential convergence to

ε-optimality. We end with a central limit theorem on the observed objective values at the sequence of generated random measures. To further intuition, we include several illustrative examples with exact influence function calculations.

5 - A Reliability Theory of Compromise Decisions for Large-Scale Stochastic Programs

Shuotao Diao, Northwestern University, Evanston, IL, United States, Suvrajeet Sen

Stochastic programming models can lead to very large-scale optimization problems for which it may be impossible to enumerate all possible scenarios. In such cases, one adopts a sampling-based solution methodology in which case the reliability of the resulting decisions may be suspect. For such instances, it is advisable to adopt methodologies that promote variance reduction. One such approach goes under a framework known as "compromise decision", which requires multiple replications of the solution procedure. We study the reliability of stochastic programming solutions resulting from the "compromise decision" process. This process is characterized by minimizing the aggregation of objective function approximations across replications, presumably conducted in parallel. We refer to the post-parallel-processing problem as the problem of "compromise decision". We quantify the reliability of compromise decisions by estimating the expectation and variance of the "pessimistic distance" of sampled instances from the set of true optimal solution set. Such pessimistic distance is defined as an estimate of the largest possible distance of the solution of the sampled instance from the "true" optimal solution set. The Rademacher average of instances is used to bound the sample complexity of the compromise decision.

MA78

Regency - 709

Simulation & Analytics for Sustainable and Resilient Urban Networks - Session 2

Invited Session

Computing Society

Chair: Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA

1 - Optimizing On-site Green Hydrogen Consumption of Heavy-duty Hydrogen Fuel Cell Electric Vehicles

Mohammad Reza Ghorbanali Zadegan, University of Central Florida, Orlando, FL, United States, Zhaomiao Guo

Transitioning to zero-emission heavy-duty freight vehicles, such as fuel cell electric vehicles (FCEVs), considerably contributes to mitigating the environmental footprint of the freight sector in transportation emissions. Our study develops a mixed-integer programming (MIP) model that optimizes the routing/scheduling, refueling, and hydrogen production operations of FCEVs, leveraging on-site green hydrogen sources. The model integrates solar-powered hydrogen production to utilize renewable resources for FCEVs refueling, and reduce reliance on non-renewable energy sources. Our findings demonstrate that this approach deals with different logistical challenges by strategically determining the locations and sizes of hydrogen refueling stations (HRSs) based on geographic and spatiotemporal conditions. Our model was initially tested using well-established Solomon benchmark instances to evaluate its robustness and efficiency. Consequently, we implemented the model in a real-world case study of freight distribution across Florida, further demonstrating its applicability. Extensive sensitivity analyses were conducted on various parameters to ensure that our model performs reliably under a variety of operational and environmental conditions. Keywords: Fuel Cell Electric Vehicles (FCEVs), Freight Transportation, Renewable Hydrogen Production, Hydrogen Refueling Stations (HRSs)

2 - Improving Resilience and Trust in AI Driven Cyber Physical Systems

Apurva Narayan, University of Western Ontario , London, ON, Canada

In today's interconnected digital era, Cyber-Physical Systems (CPS) have emerged as the backbone of critical infrastructures, encompassing a wide range of applications from smart grids to autonomous vehicles. With the advent of Artificial Intelligence (AI) technologies, these systems have witnessed unprecedented advancements, leading to improved efficiency, adaptability, and automation. However, the integration of AI introduces new challenges related to system resilience and trustworthiness, as these systems become susceptible to a new breed of threats and vulnerabilities. In this talk, he will present some of the work he's doing in his group to ensure the reliability of ML systems in the presence of adversarial attacks, primarily patch attacks in computer vision. The focus of our work has been to introduce novel strategies for practical attacks on AI driven computer vision systems and subsequently developing defences against such vulnerabilities by characterizing properties of patches. In the second part of his talk, he will showcase some recent work in the domain of digital twins where he and his group have developed novel strategies to develop robust AI systems in data scarce situations using transfer learning and active transfer learning.

3 - Integrated Modeling of Shared Mobility Systems

Hua Cai, Purdue University, West Lafayette, IN, United States, Hao Luo

Shared mobility, including bike-sharing, shared e-scooter, and ride-hailing, could improve transportation sustainability when substituting private car use and integrating with public transit. However, if shared mobility competes with other green modes, it cannot guarantee sustainability benefits. The competing and synergistic relationships between conventional modes and shared mobility are complex and not well-studied to date. We built a city-scale agent-based model that is supported by real-world mode choice survey and trip data to better study how different design and pricing of different mobility systems will lead to different competing and synergistic relationship among the shared mobility and conventional transportation systems. Using Indianapolis as the case study city, our results show that the current deployment of shared mobility systems (mainly in downtown area) does not help improve urban sustainability because the shared mobility is primarily competing with green travel modes such as walking and bus. To reduce transportation emissions, the shared micro-mobility vehicles need to be optimally deployed and rebalanced to areas that can supplement public transit network to enable multi-modal trips as an alternative to driving.

4 - Exploring the Implications of Alternative Approaches to Flood Frequency Analysis on Urban Costal Flood Attribution

Julianne Quinn, University of Virginia, Charlottesville, VA, United States, Daniel Lassiter, Daniel Wright

Characterizing flood risks in urban coastal systems is complex due to the compound, non-independent drivers of flooding (rainfall and storm surge), and nonlinear propagation of excess water through the storm sewer network. Yet understanding these risks and their drivers is critical to informing flood mitigation designs. In this study, we explore how estimates of flood volumes at different return periods and their associated drivers depend on the method used for Flood Frequency Analysis (FFA). Using the City of Norfolk, VA as a case study, we compare three FFA methods: a conventional design storm (DS) approach and two compound event (CE) approaches that use either stochastic storm transposition (SST) or stochastic storm rescaling (SSR). The DS approach simulates stylistic 24-hour precipitation time series independent of storm surge levels and assumes flood volume probabilities have the same return period as the precipitation event. The CE approaches estimate flood volume probabilities empirically from a time series of simulated storms, whose rainfall and storm surge characteristics are correlated through a copula fit to the historical record. In SST, rainfall time series are produced by stochastically transposing historical radar rainfall events in space over a meteorologically homogenous transposition domain, while in SSR, they are generated by rescaling historical storms at the same location. For each method, we decompose the flood volumes at different locations and return periods into that which can be attributed to rainfall, storm surge, and their interaction. Future work will explore the implications of these differences on optimal flood mitigation designs.

Monday, October 21, 9:45 AM - 10:35 AM

Summit - Ballroom 1

Making the most of this AI Moment: A Fireside Chat with Charles Isbell and Michael Littman

Plenary/Keynote Session

Plenary

Chair: Polly Mitchell-Guthrie, Kinaxis, Scottsdale, AZ, United States

1 - Making the most of this AI Moment: A Fireside Chat with Charles Isbell and Michael Littman

Cynthia Rudin, Duke University, Durham, NC, United States, Charles Isbell, Michael Littman

Interest in AI, which was already high, has reached a fever pitch since the release of ChatGPT two years ago. Fascination and fear abound, so how do we engage the public for the health of society to ensure AI is a force for good and not for harm? How do we broaden access to computing to shape its future? How can comedy and humor help? What roles should leading scientists take to best impact these issues? These are just a few of the questions to be discussed during this fireside chat, which longtime INFORMS member Cynthia Rudin will lead with two fellow computer scientists, who are excited to share their combined 60+ years of experience in AI and machine learning.

Monday, October 21, 10:45 AM - 12:00 PM

MB01

Summit - 320

Information Sharing in Service Operations

Invited Session

Service Science

Chair: Dongwook Shin, HKUST Business School, Clear Water Bay, Hong Kong

Co-Chair: Jeunghyun Kim, Korea University, Seoul, Korea, Republic of

1 - Recommender Systems Under Privacy Protection

Shouqiang Wang, University of Texas at Dallas, RICHARDSON, TX, United States, Can Kucukgul, Ozalp Ozer

Consumers make inferences about a product’s relevance or even get access to it through recommendations. Namely, recommendations often play both informative and allocative roles. However, the pervasive use of personal data by modern algorithmic recommender systems has sparked public outcry for tighter privacy regulations. Personal preferences over different product offerings are a basic constituent of consumer privacy. We study how a profit-driven online platform designs its recommender policies in response to different privacy protection regimes that grant users varying degrees of control over their personal data. We demonstrate the effective equivalence between the opt-out protection and the unprotected privacy. As a key finding, consumers’ autonomy over their privacy (to the extent that they could misrepresent their personal data) may compel platforms to distort their recommender policies and lead to unintended consequences. When the recommendation only plays an informative role, such level of privacy protection deters the platform from any personalized recommendation; if the recommendation can, in addition, act allocatively to control consumers’ access to products, algorithmic discrimination may arise, whereby the disadvantaged minority in the society are restricted or deprived of access to potential valuable opportunities. Ultimately, these distortions could inadvertently hurt both platforms and consumers, relative to less stringent privacy protection regimes. Counter-intuitively, enacting the recommendation’s allocative role (by restricting users’ access to certain products) in addition to its informative role can in fact benefit both the platform as well as the users, especially when users are given the autonomy over their privacy.

2 - Information Sharing to Optimize the Waiting-TIME Experience

Jeunghyun Kim, Korea University, Seoul, Korea, Republic of, Laurens Debo, Robert Shumsky

We develop a model to optimize the provision of queue-position information to waiting customers. A customer’s emotional response, and satisfaction with the wait, may be driven by loss aversion in expected waiting time and/or by disutility due to wait-time uncertainty (we call this latter effect ‘uncertainty awareness’). We consider a class of announcement strategies characterized by frequency of updates between two announcements one upon arrival and the other upon entering service. We control the frequency by the number of queue-position advancements needed to hear the next announcement and optimize it. We find that when customers are purely loss-averse, a strategy that never provides intermediate updates (arrival-only) is optimal while with purely uncertainty-aware customers, a strategy that updates customers with every single step (full information) is optimal. For customers affected by both loss aversion and uncertainty, there is an interior solution and the degree of customer satisfaction improvement (relative to the arrival-only and full information) becomes more substantial as the system becomes more loaded.

3 - The Effects of Queue-Related Information on Abandonment in The Emergency Department

Vahid Sarhangian, University of Toronto, Toronto, ON, Canada, Yaniv Ravid, Philipp Afèche, Rouba Ibrahim

Patients who leave without being seen pose a significant operational challenge to many emergency departments. We study how patients' probability of abandoning changes as they observe different events in the waiting room, by constructing a dynamic hazard model to describe patients' wait using time-dependent covariates. We find that during specific time intervals of their waiting period, observing others enter and leave the waiting area significantly affects patients' risk of abandoning before being seen by a doctor. Moreover, our results indicate that the effect changes at different congestion levels. For example, observing others arrive into the waiting room increases a patient's risk of abandoning when congestion is low, but when there are sufficiently many others waiting for service, observing others arrive reduces one's risk of abandoning. Our results can help emergency departments identify patients that are at high risk of abandoning based on the events that they have observed during their wait.

4 - Should a Platform Compete with a Third-Party Seller?

Zihao Zhang, University of Science and Technology of China, Hefei, China, People's Republic of, Ming Fan, Dongyuan Zhan

Platforms, such as Amazon, have been criticized to manipulate product information by Bayesian persuasion when competing with third-party sellers. We develop a model to analyze the phenomenon. We find that while Bayesian persuasion may improve social welfare, it causes unfair competition and always hurts third-party sellers and customers. We propose a method to reduce the strength of information manipulation and improve the fairness.

MB02

Summit - 321

Junior Faculty Network Session Across Societies

Flash Session

Quality, Statistics and Reliability

Chair: Hongyue Sun, University of Georgia, Athens, GA, United States

Co-Chair: Elynn Chen, New York University, New York, NY, 10012, United States

Co-Chair: Mansah Preko, Ghana Institute of Management and Public Administration (GIMPA), Accra, N/A, Ghana

1 - Introduction

Hongyue Sun, University of Georgia, Athens, GA, United States

. This is a network/interaction session for junior faculties, and the list of people will just briefly introduce themselves in around 2 minutes.

2 - AI/ML enabled data/knowledge fusion for quality and productivity improvements of complex engineering system

Shancong Mou is an incoming assitant Professor in the School of Industrial & Systems Engineering (ISyE) at University of Minnesota Twin Cities. He got his Ph.D. degree from ISyE, Gatech, advised by Dr. Jianjun Shi. Shancong's research interest focuses on AI/ML enabled data/knowledge fusion for quality and productivity improvements of complex engineering systems, with methodological development in robust learning, meta-learning and physics informed learning, which interfaces high-dimensional statistics, optimization research, machine learning, and computational science. His work has generated multiple papers published in top journals, recognized by various best paper awards from INFORMS QSR, IISE DAIS, and IISE QCRE divisions, and several scholarships/fellowships from ASA, IISE, ISA, and Georgia Tech. He is also the recipient of the ISyE Outstanding Graduate Student Instructor of the Year award in 2022. Shancong serves as a board member for the IISE QCRE division. He also organizes seminars, sessions, workshops, and data challenges at INFORMS, IISE, CVPR, and ECCV conferences.

3 - Balancing Optimality and Diversity: Enhancing Human Decision-Making through Generative Curation

Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

"Woody Zhu is an Assistant Professor in Heinz College of Information Systems and Public Policy, Carnegie Mellon University. He obtained his Ph.D. in Machine Learning at H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology. His research lies in the broad area of machine learning, statistics, and operations research. He is particularly interested in spatio-temporal modeling and decision making under uncertainty. Much of his work aims to develop new methodologies that address high-impact problems in various applications, including power system, renewable energy, urban planning, transportation, and healthcare. "

4 - Tensor-view Topological Graph Neural Network

Elynn Chen, New York University, New York, NY, United States

"I currently serve as an Assistant Professor in the Department of Technology, Operations, and Statistics (TOPS) at NYU Stern School of Business. Prior to this role, I was a postdoctoral fellow at the University of California, Berkeley's Department of Computer Sciences under the mentorship of Prof. Michael I. Jordan. I also held a postdoctoral position at Princeton University's ORFE, collaborating with Prof. Jianqing Fan. I received my Ph.D. degree in Statistics from Rutgers University, where I was guided by Prof. Rong Chen. My research interests are centered around developing cutting-edge methodologies for data-driven decision-making and the analysis of complex datasets, particularly in the fields of business, economics, and finance. Currently, my work is focused on advancing the areas of Tensor Learning, Reinforcement Learning, and Fusion Learning. For more information about my research and publications, please visit my website at https://elynncc.github.io/."

5 - Ethics in Action: Operationalizing Fairness in Algorithmic Decision-Making

Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

Dr. Hadis Anahideh is as an Assistant Professor of Industrial Engineering at the University of Illinois Chicago. She earned her M.Sc. and Ph.D. in Industrial Engineering and holds a B.S. in Applied Mathematics. Her research is centered on black-box optimization, active learning, machine learning, and algorithmic fairness, with a focus on developing innovative methodologies for engineering operations, design, and social systems. Dr. Anahideh is the director of the Optimal Learning and Exploration Laboratory (OPLEX) at UIC, where she leads a dynamic team of students and collaborators on projects that push the boundaries of her research interests.

6 - Advanced Machine Learning for Complex Physical Systems

Yinan Wang, Rensselaer Polytechnic Institute, Troy, NY, United States

Dr. Yinan Wang is an Assistant Professor in the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. He received the B.S. degree in Electrical Engineering and Automation from Xi’an Jiaotong University, the M.S. in Electrical Engineering from Columbia University, and the Ph.D. in Industrial and Systems Engineering from Virginia Tech. His research interests include data analytics and machine learning techniques in quality control of advanced manufacturing systems. He is the recipient of FTC Early Career Award, 10 Best Paper/Poster/Featured Article Awards, and two Best Ph.D. Dissertation Awards. He is selected as the Mary and Joseph Natrella Scholar from the American Statistical Association (ASA), Scialog Fellow from the Research Corporation for Science Advancement (RCSA), and the Panel Fellow in the NSF CMMI Game Changer Academy. He was co-organizer for the Symposium and Workshop in Manufacturing Science and Engineering Conference (MSEC) 2023 and SIAM International Conference on Data Mining (SDM) 2023 and 2024. He serves as an active Board Director in the IISE Data Analytics and Information Systems (DAIS) division and program director in the Automatic Controls and Robotics Division (ACARD) of the International Society of Automation (ISA).

7 - Structured Data Analysis with Weak Supervision in Healthcare Applications

Xin Zan, University of Iowa, Coralville, IA, United States

Xin Zan is joining the University of Iowa as an assistant professor in the Department of Industrial and Systems Engineering in Fall 2024. Xin received her BS degree in industrial engineering from Shanghai Jiao Tong University, China, in 2019, the master’s degree in statistics and Ph.D. degree in industrial and systems engineering from the University of Florida in 2022 and 2024, respectively. Her research interests are focused on advanced data analytics for complex system modeling, monitoring, and diagnosis to improve system performance with applications mainly in healthcare. Xin’s research works have had a broad impact and been recognized by several distinguished awards from INFORMS, IISE, and COVID Information Commons (CIC), etc. She is a member of INFORMS and IISE.

8 - Toward Explainable Time Series Data Mining via Motif

Yifeng Gao, University of Texas Rio Grande Valley, Brownsville , TX, United States

Researchers work on time series data driven health applications

9 - Privacy-aware Time Series Data Sharing

Li Zhang, University of Texas Rio Grande Valley, Brownsville, TX, United States

Li Zhang is an assistant professor in the Department of Computer Science at the University of Texas Rio Grande Valley (UTRGV). She received his Ph.D. degree at George Mason University in 2023. She has worked on time series data mining for five years and published in premier data mining conferences in data mining research fields such as ICDM and SDM. She is currently interested in developing robustness and interpretable time series models for anomaly detection, forecasting, evolving pattern discovery methods in manufacture, health care and renewable energy related applications.

10 - Process Monitoring and Control of Advanced Manufacturing based on Physics-Assisted Machine Learning

I am an Assistant Professor in the Department of Industrial Engineering at Pusan National University in Korea. I received a Ph.D. in Industrial and Systems Engineering at Virginia Polytechnic Institute and State University (May 2023). My research area is developing data-driven methods, including artificial intelligence, machine learning, and statistical learning, to achieve quality assurance in advanced manufacturing processes and solve various problems in healthcare systems and material analysis.

11 - Process Monitoring and Control of Advanced Manufacturing based on Physics-Assisted Machine Learning

Jihoon Chung, Virginia Tech, Blacksburg, VA, United States

I am an Assistant Professor in the Department of Industrial Engineering at Pusan National University in Korea. I received a Ph.D. in Industrial and Systems Engineering at Virginia Polytechnic Institute and State University (May 2023). My research area is developing data-driven methods, including artificial intelligence, machine learning, and statistical learning, to achieve quality assurance in advanced manufacturing processes and solve various problems in healthcare systems and material analysis.

12 - The Sooner, the Better? Optimal Vaccination Policy with Limited Vaccine Supply

Miao Bai, University of Connecticut, Storrs, CT, United States

Miao Bai is an assistant professor in the Department of Operations and Information Management at the University of Connecticut, Storrs, CT and a research collaborator at Mayo Clinic, Rochester, MN. His primary research interest is to analyze and address complex problems in the domains of public health, healthcare operations management, and medicine.

13 - Recent advances on multi-task learning and fine-tuning: An algorithmic perspective

Hongyang Zhang, Northeastern University, Boston, MA, United States

Recent advances on multi-task learning and fine-tuning: An algorithmic perspective

14 - "Empowering Frontline Health Workers to Tackle Stock-outs in LMICs: Evaluating Training Programs in Indonesia"

Amir Karimi, The University of Texas at San Antonio, San Antonio, TX, United States

"Professor Amir Karimi's research is positioned at the intersection of Global Health, Healthcare Supply Chains, and Socially Responsible Supply Chains. In particular, his research endeavors are focused on improving access to essential health commodities (e.g., contraceptives, antimalarials, HIV medication) in Low- and Middle-income Countries (LMICs). For individuals who are deprived access to such essential health commodities , the consequences can be dire. For example, without reliable access to contraceptives, women may suffer unintended pregnancies, imposing economic and psychological burden, and adverse health outcomes including injuries, infections, and even deaths. HIV patients unable to obtain antiretrovirals may face delays in treatment initiation and interruptions during the treatment process, subsequently increasing the risk of viral resistance, treatment failure, and mortality. Focusing on this context, Amir's research aims to: (i) Empirically evaluate and uncover the factors that impact health commodity access in LMICs by leveraging field data and using a combination of rigorous econometric and predictive modeling techniques; (ii) Generate actionable insights that public health organizations, governments, and donors can use to improve health commodity access in LMICs. Amir has a Ph.D. in Supply Chain and Operations from the University of Minnesota, a master's degree in Production and Operations Management from University of Tehran, and a bachelor's degree in Industrial Management from Shiraz University."

15 - New Wine in Old Bottle: Statistical Learning in Operations Management

Zhengyuan Zhou, Stern School of Business, New York University, New York, NY, United States

Zhengyuan Zhou is currently an assistant professor in New York University Stern School of Business, Department of Technology, Operations and Statistics. Before joining NYU Stern, Professor Zhou spent the year 2019-2020 as a Goldstine research fellow at IBM research. He received his BA in Mathematics and BS in Electrical Engineering and Computer Sciences, both from UC Berkeley, and subsequently a PhD in Electrical Engineering from Stanford University in 2019.

16 - Generative AI in Action: Applications in Consumer Insights, Mental Health, and More

Yuan Zhang, University of Memphis, Memphis, TN, United States

Dr. Yuan Zhang is an assistant professor in Management Information Systems in the Fogelman College of Business and Economics at the University of Memphis. Mainly using applied econometrics, experiments, and causal inference machine learning, her research investigates user experience, user behaviors, and IT artifacts design focusing on usability and affordance primarily in the contexts of social media and video games. Her research has been published in major information systems and interdisciplinary journals including JAIS and I&M and presented at major conferences including CIST, INFORMS, WISE, HICSS, WITS, ACR.

17 - Robustness in Data-Driven Decision-Making

Mohammed Amine Bennouna, Massachusetts Institute of Technology, Cambridge, MA, United States

Amine Bennouna is an incoming Assistant professor at Kellogg School of Management at Northwestern University. He completed his PhD at MIT's Operations Research Center. His research interests lie at the intersection of machine learning, optimization and statistics, with a focus on developing novel machine learning and data-driven decision-making algorithms with desirable properties such as robustness and interpretability. Prior to joining MIT, he received a Bachelor and Master of Science in Applied Mathematics from Ecole Polytechnique.

18 - Active Learning Techniques for Uncertainty Quantification

Ozge Surer, Miami University, Oxford, OH, United States

Özge Sürer is an Assistant Professor of Business Analytics at Miami University. Before joining Miami University, she was a postdoctoral research fellow at the Northwestern Argonne Institute of Science and Engineering, where she specialized in developing novel techniques utilizing Bayesian uncertainty quantification and computational statistics. Dr. Sürer earned her Ph.D. in Industrial Engineering and Management Sciences from Northwestern University in December 2020, with her dissertation centered on creating interpretable predictive models for large data sets.

19 - On algorithms and complexity to address high-dimensional nonconvex optimization

Yue Xie, University of Hong Kong, Pokfulam, Hong Kong

Dr. Yue Xie is a Research Assistant Professor in Musketeers Foundation Institute of Data Science (HKU-IDS) and Department of Mathematics at the University of Hong Kong. He was a postdoc at UW Madison working in the nonconvex optimization group led by Professor Stephen J. Wright. He received his PhD degree in Pennsylvania State University and Bachelor degree from Tsinghua University. Dr. Yue Xie has been focusing on algorithm design and analysis to address large-scale nonconvex and stochastic optimization problems with all types of applications including machine learning and data science. He has published/served as the referee of top-tier journals including Mathematical Programming, SIAM Journal on Optimization, and IEEE Transactions on Automatic Control, etc. More details about him can be found at: https://yue-xie.github.io.

20 - Data-Driven Methods for Healthcare Operations

Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

Savannah Tang is an Assistant Professor at SMU Cox School of Business. She earned her Ph.D. in Operations Management from the Tepper School of Business at Carnegie Mellon University and also holds a Master's degree in Machine Learning from CMU's School of Computer Science. Savannah's research focuses on data-driven decision-making in critical real-world applications, including organ allocation, child welfare, and personalization.

21 - AI in Healthcare - Are African Healthcare Policies Ready?

Mansah Preko, Ghana Institute of Management and Public Administration , Accra, Ghana

Dr. Mansah Preko is an information systems researcher, an industry IT professional, and an adjunct lecturer at the Ghana Institute of Management and Public Administration (GIMPA). She holds a Ph.D. in Information Systems (IS) with specialization in digital health from the University of Ghana, a Master of Science in Management Information Systems from Coventry University, and a Bachelor of Science in Information and Communication Technology from GIMPA. Her studies have covered a wide range of topics in the areas of digital transformation, business intelligence, eHealth, and information systems in general. Over the years, Dr. Preko has addressed some of the pertinent issues in the field of IS and developing economies through data-driven research. These include her works that focused on the role of technology in promoting doctor entrepreneurship, the role of technology in mitigating brain drain in the health sector, the challenges and mitigating strategies in digitalizing healthcare, the unintended negative consequences of digitalization in the health sector, stakeholder conflicts in healthcare digitalization, the role of business intelligence in the banking industry, the use of drones in transforming the healthcare industry, among others. Her interest is also found in 'neoliberalism and gender' in higher education, as well as women in STEM research. Dr. Preko serves as an editorial reviewer and advisory board member for several IS and engineering journals and conferences across the globe.

She is a certified Project Management Professional (PMP) and a member of various professional associations. Dr. Preko can be reached at mansah.preko@gmail.com

22 - Score2Text: a Large Language Model for Interpretable Credit Scoring

Huan Yu, University of Southampton, Southampton, United Kingdom

Dr Huan Yu is a Lecturer in Business Analytics within Southampton Business School at the University of Southampton. Huan is also a member in CORMSIS (Centre for Operational Research, Management Science and Information System) at the University of Southampton. Huan joined the Southampton Business School as a Lecturer in September 2019. Before Huan joined the University of Southampton, she completed her PhD in Management Science and Engineering from the School of Management at the University of Science and Technology of China in May 2019. During her PhD, she was also a Visiting Researcher at IESEG School of Management in 2014-2016 and National University of Singapore in 2018.

23 - From Full-Time to Flexi: Examining the Outcomes of Workforce Changes in Primary Healthcare Delivery

Harshita Kajaria-Montag, Indiana University, Bloomington, IN, United States

From Full-Time to Flexi: Examining the Outcomes of Workforce Changes in Primary Healthcare Delivery

24 - An Efficient Algorithm for Large-Scale Stochastic Optimization with Constraints

Baoyu Zhou, Arizona State University, Tempe, AZ, United States

Baoyu Zhou is an Assistant Professor in the School of Computing and Augmented Intelligence at Arizona State University (ASU). Before joining ASU, he was a postdoctoral researcher at the University of Michigan and the University of Chicago. He received his Ph.D. in Industrial and Systems Engineering from Lehigh University. He won the Elizabeth V. Stout Dissertation Award at Lehigh University in 2022. His research focuses on developing, analyzing, and implementing efficient algorithms for solving large-scale optimization problems that are potentially nonlinear, stochastic, and constrained.

25 - Structural properties of integer programming with bounded subdeterminants

Luze Xu, University of California, Davis, Davis, CA, United States

"Luze Xu currently holds the position of Krener Assistant Professor in the Department of Mathematics at UC Davis, mentored by Professor Matthias Köppe and Professor Jesús De Loera. He received his Ph.D. in Industrial and Operations Engineering at the University of Michigan in April 2022, under the guidance of Professor Jon Lee. He is an optimizer specializing in mixed integer programming. His research interests include integer programming, mixed integer programming, global optimization, and theories that bridge continuous and discrete optimization. His current research focuses on structural integer programming with bounded subdeterminants, mixed integer semidefinite optimization, and mixed integer quadratic optimization."

MB03

Summit - 322

Data Analytics and Machine Learning for Advanced Manufacturing processes

Invited Session

Quality, Statistics and Reliability

Chair: Zimo Wang, SUNY Binghamton, Binghamton, NY, United States

Co-Chair: Chen Kan, The University of Texas at Arlington, Arlington, TX, United States

1 - Debiasing Machine Learning with Partial Information for Design and Manufacturing

Yanwen Xu, University of Texas at Dallas, Richardson, TX, United States

Uncertainty quantification stands as a foundational element in both design and manufacturing processes. However, accurately quantify uncertainty within complex engineering systems presents a formidable hurdle due to the involvement of numerous parameters, thereby imposing substantial computational demands for failure testing. To mitigate the computational costs associated with uncertainty quantification, researchers have turned to machine learning models as substitutes for expensive physical models. We are particularly interested in the Gaussian process (GP) model, which has garnered increasing attention in manufacturing applications. Constructing GP-based machine learning models typically necessitates fully observed information or sample points. However, many engineering scenarios involve missing values in collected datasets, stemming from factors such as unobserved input features, unreliable sensors, and data collection errors. The availability of only limited fully observed data results in increased variance and diminished model accuracy. ***Integrating partially observed datasets could potentially expand the dataset size, but naively imputing the missing data may introduce bias.*** To address the challenge posed by partially observed information, ***this paper introduces a hierarchical framework designed to alleviate bias in machine learning***. The framework aims to leverage all available information comprehensively, rather than solely relying on the fully observed portion. We provide theoretical guarantees (asymptotic normality) for our approach, which can be used for inference and construct confidence intervals. This is pivotal for design and manufacturing since it furnishes a spectrum of plausible parameter values, providing valuable insights into the precision and uncertainty of the estimate.

2 - A Shapley explanation-based active learning machine for inferencing high dimensional acoustic informatics: an application for in-situ nanomachining monitoring

Xinchen Wang, Binghamton University, Vestal, NY, United States, Zimo Wang

Understanding streaming sensor signals is critical for in-process monitoring and quality inspection for various industry applications. The streaming sensor signals are induced with real-time information of the underlying processes, allowing immediate anomaly detection and diagnosis for potential quality issues. Critical issues arise as the monitored process scale decreases to micro- and even nano-level. Due to the unavoidable acquisition step, the obtained information is overwhelmingly submerged in high-level noises, which may adversely impact the accuracy and effectiveness of extracting information from monitored sensor signals. In addition, without understanding the sources and frequency responses of the environmental noises under high precision level, it is impractical to directly screen out the interferences using conventional techniques (e.g., band filtering and denoising autoencoder). This paper proposes an active learning strategy for the streaming signal data with the shapely embedded space projection from high dimension Fourier domain to low dimension spectral domain to allow the recognition of weak signals in high noise level environments. The proposed approach is tested by experimental music timbre recognition problems followed by real-time surface morphology characterizations under nanomachining processes. The results indicate that the proposed approach can effectively capture the underlying spectral patterns of acoustics signals. Further investigations into the sensor-based monitoring scheme for the nanomachining process suggest that the presented approach achieves an accuracy of around 92% for predicting the cutting depth and 82% for cutting width in real-time. The presented framework opens up an avenue to allow real-time characterizations and quality inspection for surface characteristics under the nanoscale.

3 - Incremental Profiling of 3D Point Clouds for In-Situ Quality Assurance of Additive Manufacturing

Chen Kan, The University of Texas at Arlington, Arlington, TX, United States, Zehao Ye

In this work, we introduce an innovative framework for point cloud-based quality control of additive manufacturing (AM). Point clouds have shown great potential in AM process monitoring and anomaly detection. However, given the high-volume, high-dimensionality, and unstructured nature of point clouds, it is critically challenging to efficiently process them and extract defect-pertinent information, especially in a real-time manner. To address this challenge, this work develops an efficient point cloud representation learning approach that updates the low-dimensional geometric profile of the point cloud in an incremental way. This is particularly suited for layer-wise monitoring and detection of geometric defects in AM builds. The developed framework is evaluated and validated in simulation and real-world studies, which demonstrate its effectiveness.

4 - Mixed Effects Model via Distribution-in-distribution-out Regression

Mengfan Fu, University at Buffalo, Buffalo, NY, United States, Yujing Huang, Xiaoyu Chen, Xinwei Deng

Mixed effects model has been the key technique to jointly consider fixed effects and random effects in the analysis of an experiment. It requires one to first create a design of experiment by generating design matrices for both fixed effects and random effects. Motivated by the imperfect controllers in many manufacturing and healthcare systems, rather than considering deterministic design matrices, we propose to create such design matrices as probability distributions. These distributional design matrices allow one to consider the randomness of the controllers that realize the design, leading to distributional response. To analyze the deterministic relationship between distributional design matrices and the distributional responses, we propose to employ a distribution-in-distribution-out (DIDO) regression model by analyzing such relationship on 2-Wasserstein space. We will also demonstrate that the existing mixed effects model is a special case of such DIDO regression model when reducing to Euclidean space. Simulation studies and a case study were created to validate the proposed methodology.

MB04

Summit - 323

Modeling Complex and High-dimensional Data for Healthcare Engineering

Invited Session

Quality, Statistics and Reliability

Chair: Manni Zhang, Purdue University, West Lafayette, IN, United States

Co-Chair: Ana Estrada Gomez, Purdue University, West Lafayette, IN, United States

1 - missRASS-CPD: Variable Selection for Multimodal Data from Multiple Groups with Missing Values

Manni Zhang, Purdue University, West Lafayette, IN, United States, Ana Estrada Gomez

Nowadays, multi-modal data collected from complex systems is ubiquitous. In such systems, each modality provides myopic information about the system. Therefore, fusing the data from all the modalities is critical to understanding the system as a whole. For example, to diagnose Inflammatory Bowel Diseases, ulcerative colitis (UC) and Crohn's disease (CD), different omics data are needed. Transcriptomic data provides information about gene expressions, proteomic data about protein levels, while metabolite data provides information about microorganisms' concentrations. Using only one type of omics data results in limited diagnosis accuracy and poor treatment plans. Furthermore, to improve diagnosis and treatment, differences and similarities across groups (UC, CD, and healthy individuals) need to be incorporated into the analysis. In this work, we propose a framework that fuses multi-modal data while exploiting the similarities and differences across groups for accurate diagnosis. The framework finds the joint latent variables for each group and the individual latent variables for each modality, and identifies the key features in each modality. Furthermore, the framework is capable of imputing missing data and removing outliers. Our method decomposes the data into a low-rank component, that captures the true signal, and a sparse component, that captures the outliers. Then, using CP decomposition on the low-rank component with a sparse group Lasso penalty, we simultaneously learn the joint and individual latent variables. The performance of the proposed framework is illustrated through simulated and real data examples. In particular, we use the proposed framework to identify key biomarkers for UC and CD patients.

2 - Multi-Modal Graph Neural Networks for Predictive Modeling of Multiple Chronic Conditions

Adel Alaeddini, The University of Texas at San Antonio (UTSA), San Antonio, TX, United States, Julian Carvajal Rico, Joseph Mccormick, Susan Fisher-Hoch

We developed a specialized Cross-Modality Graph Neural Network (GNN) to learn the stochastic relationship between patient's pre-existing conditions and future chronic conditions based on their metabolic information. The model was validated using a large longitudinal cohort study.

3 - Functional Connectome Fingerprinting through Semi-symmetric Tucker2 Tensor Decomposition

Vitor Farias Costa de Carvalho, Purdue University, West Lafayette, IN, United States, Mintao Liu, Ana María Estrada Gómez, Joaquín Goñi

The human functional connectome (FC) represents the functional couplings between brain regions derived from blood oxygen level-dependent (BOLD) signals. Studying the FC enables obtaining insights about individual traits from subjects. In this study, we aim to maximize FC identifiability by performing semi-symmetric Tucker decomposition in data from the Human Connectome Project (HCP). Identifiability is the general paradigm of correctly pairing FCs that belong to the same subject within a dataset where there is test and retest for all subjects. Here we measure fingerprinting via a measurement denominated matching rate. Studying identifiability through FCs enables verifying their consistency over time and understanding how subjects differ from each other at an individual level. Here we used test-retest fMRI data from 426 unrelated subjects from the Young Adult HCP dataset. For each subject, test-retest data includes resting state and seven tasks (Gambling, Emotion, Language, Motor, Relational, Working Memory, and Social). To analyze the high-complexity data, we structured it into a tensor and used tensor decomposition to project the data into a lower dimensional space. By doing so, the first two components capture information related to the brain's functional couplings, whereas the third component captures subject-specific information. We repeated these analyses for three brain parcellations. Results show that subjects can be correctly matched with 71-100% accuracy across resting state and all evaluated tasks. Higher granularity led to higher matching rates for all fMRI conditions. We also extended this approach to assess fingerprinting when using resting state data as test and task data as retest.

MB05

Summit - 324

DEI Special Issue of Production and Operations Management: Selected Papers

Invited Session

Diversity, Equity, and Inclusion

Chair: Michael Johnson, University of Massachusetts Boston, Boston, MA, 02125, United States

1 - No Country for Young Refugees: Barriers and Opportunities for Inclusive Refugee Education Practices

Sebnem Manolya Demir, Marshall School of Business, University of Southern California, Los Angeles, CA, United States, Feyza Guliz Sahinyazan, Bahar Y. Kara, Elfe Buluc

The recent refugee crises in Ukraine and Syria have created millions of refugees, 40% of whom are children. The education systems of countries hosting refugees struggle to integrate such large populations. In addition, language barriers and stigma hamper inclusive and equitable education opportunities for refugee children. “Lost generations” distanced from education may have a long-term dependency on social security systems and monetary aid. This study considers the following research question: How can a host country improve the inclusion of refugee children in the education system without overburdening its infrastructure? First, we document the availability and accessibility challenges and opportunities that refugee children face during the Syrian refugee crisis. We then develop an inclusive planning strategy aligned with existing capacity and resources and formulate two adaptations of the maximum covering problem (MCP): cooperative capacitated MCP with heterogeneity constraints (CCMCP-HC) to improve the current schooling access in Türkiye and Modular CCMCP-HC to guide early planning in the case of a future crisis. Our computational analyses illustrate that the proposed approach yields higher schooling rates and capacity utilization than existing approaches. Our results emphasize the importance of having a planning strategy in the initial phases of a crisis that considers future integration possibilities.

2 - Toward a More Diverse and Equitable Food Distribution System: Amplifying Diversity, Equity, and Inclusion in Food Bank Operations

Mikaya Hamilton, North Carolina A&T State University, Greensboro, NC, United States, Benjamin Morrow Jr., Lauren Davis, Shona Morgan, Julie Ivy, Steven Jiang, Min Chi, Kyle Hilliard

This article provides an evidence-based discussion of an ongoing effort within the operations of hunger relief organizations to address diversity, equity, and inclusion (DEI) by sourcing and distributing more culturally relevant food. Through nearly 100 interviews with food bank personnel in diverse roles (from partner agency relations to executives) representing various regions of the United States, we explore the challenges faced by organizations' different functional units. These interviews indicate a shift to more inclusive language, more personalized metrics, and more inclusive operations. We critically analyze the related literature and identify opportunities for infusing DEI practices in studying hunger relief supply chains.

3 - Diversity in Frontline Employee Perceptions: Policies and Procedures, Training, and Leadership as Drivers of Service Equality

Ken Kelley, University of Notre Dame, Notre Dame, IN, United States, Eve Rosenzweig, Elliot Bendoly

Excellent service assumes equivalency in its application, yet reality is far more complex. In hotel settings, providing excellent service to diverse guests is nuanced when frontline employees themselves are diverse. While operational tactics are used to promote excellent service, it is unclear whether training, policies and procedures, and leadership designed to advance excellence have the same impact on employees who, by virtue of their own background, are more attuned to guests' needs. We show that operational tactics impact frontline employees' perceptions of service equality, with racial/ethnic minority employees perceiving different levels. Employing a sample of 25,698 employee-year observations across 32 luxury hotels in the US over 3 years, we find that training and codified policies and procedures improve perceptions of guest service equality. However, racial/ethnic minority employees are less impacted than their White counterparts by leadership stances that seem to promote equality more broadly. After controlling for year and other employee- and hotel-level variables, operational tactics (a) improve perceptions of service equality and (b) reduce the disparity between White and racial/ethnic-minority service-quality assessments. Our findings demonstrate that managers can elevate perceptions of customer service equality with training and policies and procedures, thus making operations management a key driver of equality.

MB06

Summit - 325

Meet the Editors Panel: Supply Chain & Logistics, IISE Transactions

Panel Session

Location Analysis

Co-Chair: Jennifer Ryan, University of Nebraska - Lincoln, Lincoln, NE, 68588, United States

1 - Panelist

Rajan Batta, University at Buffalo, Buffalo, NY, United States

2 - Panelist

Marc Posner, The Ohio State University, Columbus, OH, United States

3 - Panelist

Larry Snyder, Lehigh University, Bethlehem, PA, United States

4 - Panelist

Mehmet Gumus, McGill University, Montreal, QC, Canada

MB07

Summit - 327

Infrastructure Planning to Address Climate Change

Flash Session

Chair: Alireza Azadnia, George Mason University, 4400 University Drive, Fairfax, VA, 22030, United States

1 - Assessing and Reducing the Environmental Impacts of Global Maritime Freight Flows in Changing Climate Conditions

Alireza Azadnia, George Mason University, Fairfax, VA, United States, Elise Miller-Hooks

This talk presents an optimization-based approach to estimating greenhouse gas (GHG) emissions and other air pollutants emitted by cargo ships in the maritime system. Based on the relationships and criteria presented in the Fourth IMO GHG Study 2020 for vessel emissions by fuel and engine type, meteorological conditions, vessel speed, and vessel draught, the model builds on a previously developed mixed-integer linear programming formulation, Benders branch-and-cut solution methodology and global maritime network representation to estimate emissions produced by global vessel traffic under changing climate conditions. The model is extended to incorporate optimal speed and routing decisions that can aid in meeting global climate targets.

2 - Research on the mechanism of regional innovation ecosystems, including 31 provinces and cities

Chengzhao Li, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

The fundamental component in comprehensively executing the national innovation-driven development strategy is regional innovation capability. It is crucial to identify the factors that influence this capability and to investigate the mechanism that links regional innovation capabilities to expedite regional innovation capability enhancement. Establishing innovation ecosystems (IEs) in different provinces and regions poses a significant challenge for enterprises seeking to thrive in the aftermath of the pandemic. This study focuses on the IEs of 31 provinces and cities, examining the impact of the regional ecosystem configuration on regional innovation performance across various dimensions through fuzzy-set qualitative comparative analysis (fsQCA).

Two high-innovation configurations are identified, namely, creative acquisition and knowledge co-creation, as well as two non-high-innovation configurations, namely, environmental alienation and technology block. The findings highlight the developmental trajectory of regional innovation ecosystems (RIEs) and underscore the significance of implementing supportive financing and talent development policies within a region.

This study employs configuration analysis to examine various antecedent conditions that can enhance innovation performance in enterprises. It focuses on the post-epidemic conditions in China and establishes a comprehensive research framework centered on RIEs.

3 - Designing and Optimization of Sustainable Supply Chain for Converting Agricultural Waste to Bioenergy

Maryam Roudneshin, University College Dublin, Dublin, Ireland, Amanda Sosa

The growing concerns about global warming and the need for energy security have underscored the importance of exploring alternative energy sources to fossil fuels. The use of agricultural waste for bioenergy production has attracted significant attention for its ability to reduce environmental impacts while creating sustainable energy sources. However, achieving profitability and attractiveness to stakeholders and investors requires optimal supply chain network design, including the selection of appropriate locations for plants. Thus, this study sets out to develop a supply chain network design that prioritizes sustainability (economic, environmental, and social considerations). Initially, a two-pronged approach was employed to identify potential biorefinery locations, leveraging both Geographic Information System (GIS) and

MCDM to balance environmental and economic factors. Subsequently, an innovative mathematical model was presented to determine the optimal bioenergy supply chain configuration in Ireland.

4 - Strategic Performance Indicators for Building Resilient Supply Chains

Afrooz Moatari-Kazerouni, Widener University, Chester, PA, United States, Yvonne Lederer Antonucci

This study undertakes an extensive examination of operational-level Key Performance Indicators (KPIs), which are crucial for enhancing the resilience of supply chains. As a pivotal element in supply chain management, resilience refers to the ability of an organization to quickly recover from disruptions while maintaining high standards of service quality and ensuring customer satisfaction. This research draws upon a broad spectrum of established theories and existing literature to delineate eight core capabilities that embody supply chain resilience: Anticipation, Sustainability, Visibility, Flexibility, Agility, Adaptability, Strategic Alignment, and Collaboration. Through a comprehensive literature review, the study identifies specific KPIs associated with each resilience capability, offering organizations practical metrics for assessing their supply chain performance. The selection of KPIs is geared towards enabling companies to effectively anticipate disruptions, adapt to changes, and recover from setbacks while fostering sustainability and strategic collaboration. To validate these findings, expert opinions were collected via a structured survey, which confirmed the relevance and applicability of the identified KPIs. This research contributes to the field by providing a detailed, actionable framework that practitioners and scholars can utilize to measure and enhance supply chain resilience. The outcomes aim to equip businesses with the tools needed to sustain competitive advantage and thrive in dynamic market environments.

MB08

Summit - 328

OR in Defense Planning

Invited Session

Military and Security

Chair: Isabella Sanders, Dept. of Systems Engineering, United States Military Academy, Newburgh, United States

1 - Microgrid capacity planning

Daniel Reich, Naval Postgraduate School, Monterey, CA, United States

We present a heuristic search method for distrubuted energy resource sizing, released in Microgrid Planner, an open-source software platform. Our method is constructed to identify a wide range of microgrid design options that satisfy a given set of power load requirements, allowing a decision maker to weigh trade-offs between potential designs and select preferred solutions. We introduce a global binary search algorithm to build a diverse set of microgrid design options and refine them using a local linear search method.

2 - The Maximal Covering Location Disruption Problem

Brian Lunday, Air Force Institute of Technology, WPAFB, OH, United States

This research sets forth and examines a new sequential, competitive location problem. The *maximal covering location disruption problem* is a zero-sum Stackelberg game comprised of two stages. A leader denies access to at most q out of n possible facility locations in the first stage and, in the second stage, a follower solves a maximal covering location problem while emplacing at most p facilities. Identifying this problem as both relevant and unaddressed in the current literature, this research examines properties of the bilevel programming formulation to inform heuristic development, subsequently evaluating the efficacy and efficiency of two variants each of an iterative, bounding heuristic (IBH) and a reformulation-based construction heuristic (RCH) over a two sets collectively consisting of 2160 test instances representing a breadth of relative parametric values. Although we illustrate that each heuristic may not identify an optimal solution, computational testing demonstrates the superlative and generally excellent performance of the RCH variants. For the 12.4% of instances for which the RCH does not readily verify the optimality of its solution, lower-bounding procedures characterize solution quality. Both of the RCH variants attain solutions with an average 4.08% relative optimality gap, and they scaled well over different parametric value combinations, solving instances in an average of 98.0 and 123.6 seconds, respectively.

3 - Impact Assessment and Course of Action Generation with Reinforcement Learning

Matthew Powers, MITRE, McLean, VA, United States

These MITRE models leverage reinforcement learning to yield high-quality COAs and COA utility scores. The contested logistics model produces strategies in operational environments requiring maneuver from strategic locations. The SOCEUR-sponsored SOF Impact Model assesses non-kinetic battlefield impact. These models propose RL and human-machine interface to overcome human cognitive limitations.

4 - Quantifying Capability Gaps via Information Relaxation and Deep Reinforcement Learning in Infinite-Horizon Markov Decision Processes: A Military Air Battle Management Application

Joseph Liles, Air Force Institute of Technology, WPAFB, OH, United States, Matthew Robbins, Brian Lunday

In this paper, we investigate the application of an information relaxation technique for approximating upper bounds on solution quality to a complex stochastic and dynamic assignment problem in military air battle management. Information relaxation refers to relaxing the non-anticipativity constraints in a sequential decision-making problem that require a decision-maker to act only on currently available information. We implement a temporal event horizon that provides a decision-maker with adjustable access to the outcome of future stochastic uncertainties in the problem environment. Whereas previous work has investigated information relaxation with regard to problems that can be solved more easily under a deterministic relaxation, we demonstrate a methodology for applying the approach to a continuous-time problem in continuous space that requires a solution approximation technique even in deterministic conditions. We formulate the problem as a discounted, infinite-horizon Markov decision process and solve it by employing a deep neural network-based approximate policy iteration algorithm in concert with several designed computational experiments. We demonstrate how a multidimensional sensitivity analysis of the event horizon and other problem features helps quantify potential improvements to decision policy effectiveness resulting from either a change to tactics or a modification to capabilities. Applied to real-world capability gap assessments, the quantification objectively augments traditionally subjective analysis to guide decision-making and establish more effective requirements for acquisition programs.

MB09

Summit - 329

Algorithms, Inference, and Incentives

Invited Session

Auctions and Market Design

Chair: Alireza Fallah, UC Berkeley, Berkeley, CA, United States

Co-Chair: Michael Jordan, UC Berkeley, Berkeley, CA, United States

1 - The Limits of Price Discrimination Under Privacy Constraints

Alireza Fallah, UC Berkeley, Berkeley, CA, United States, Michael Jordan, Ali Makhdoumi, Azarakhsh Malekian

We study a producer's problem of selling a product to a continuum of privacy-conscious consumers, where the producer can implement third-degree price discrimination, offering different prices to different market segments. We consider a privacy mechanism that provides a degree of protection by probabilistically masking each market segment. We establish that the resultant set of all consumer-producer utilities forms a convex polygon, characterized explicitly as a linear mapping of a certain high-dimensional convex polytope into $\mathbb{R}^2$. This characterization enables us to investigate the impact of the privacy mechanism on both producer and consumer utilities. In particular, we establish that the privacy constraint always hurts the producer by reducing both the maximum and minimum utility achievable. From the consumer's perspective, although the privacy mechanism ensures an increase in the minimum utility compared to the non-private scenario, interestingly, it may reduce the maximum utility. Finally, we demonstrate that increasing the privacy level does not necessarily intensify these effects. For instance, the maximum utility for the producer or the minimum utility for the consumer may exhibit nonmonotonic behavior in response to an increase of the privacy level.

2 - Leveraging Inspections and Penalties in Mechanism Design

Xuanjie Li, Duke University, Durham, NC, United States, Ali Makhdoumi, Alexandre Belloni

We consider a mechanism design problem where a seller allocates multiple items to multiunit-demand buyers. The seller can inspect the buyers to learn their types and use this information to reward truthful buyers through payments. We characterize the (approximately) optimal mechanism in this setting. We first characterize the solution to the single-buyer problem that constitutes selling a single item to a single buyer subject to an upper bound on the allocation probability. The optimal allocation strategy of the single-buyer problem is described as a continuous and monotonically increasing function of the types and can be attained by solving an ordinary differential equation. With the solution to the single-buyer problem, and by using connections to the magician problem, we design a mechanism that achieves $(1-\sqrt{\alpha})$ of the optimal revenue of the general problem, where each buyer requires at most α fraction of all units of one item. Our mechanism can be extended to the case where the underlying value distribution is erroneous but has a small total variation distance from the true distribution.

3 - Debiasing Switchback Price Experiments with Forward-Looking Demand

Yifan Wu, Northwestern University, Evanston, IL, United States, Ramesh Johari, Orrie Page, Vasilis Syrgkanis, Gabriel Weintraub

We consider a retail seller who is running pricing experiments over time, for a single product, for a sufficiently long-horizon. At each period, the seller chooses a price p from a set of predefined prices, with some probability distribution over these prices. The goal is to estimate the demand and the gradient of the demand around a particular price point, with the goal of incrementally adjusting price to improve revenue after the experiment. Crucially, buyers are forward looking with a discounted utility and will choose to not purchase now, if they expect to face a lower price in the near future, due to experimentation. We show that with a simple three-price level experiment, the seller can debias the demand and gradient estimates and recover the true demand and gradient, that removes the bias due to strategic price-anticipatory behavior.

4 - Reduced Rank Multi-Objective Policy Learning and Optimization

Ezinne Nwankwo, University of California, Berkeley, Berkeley, CA, United States, Michael Jordan, Angela Zhou

Evaluating the causal impacts of possible interventions is crucial for informing decision-making, especially towards improving access to opportunity. However, if causal effects are heterogeneous and predictable from covariates, personalized treatment decisions can improve individual outcomes and contribute to both efficiency and equity. In practice, however, causal researchers do not have a single outcome in mind a priori and often collect multiple outcomes of interest that are noisy estimates of the true target of interest. For example, in government-assisted social benefit programs, policymakers collect many outcomes to understand the multidimensional nature of poverty. The ultimate goal is to learn an optimal treatment policy that in some sense maximizes multiple outcomes simultaneously. To address such issues, we present a data-driven dimensionality-reduction methodology for multiple outcomes in the context of optimal policy learning with multiple objectives. We learn a low-dimensional representation of the true outcome from the observed outcomes using reduced rank regression. We develop a suite of estimates that use the model to denoise observed outcomes, including commonly-used index weightings. These methods improve estimation error in policy evaluation and optimization, including on a case study of real-world cash transfer and social intervention data. Reducing the variance of noisy social outcomes can improve the performance of algorithmic allocations.

5 - Techniques for Trajectory-Constrained Optimization

Jad Salem, US Naval Academy, Annapolis, MD, United States, Swati Gupta, Vijay Kamble

Exploratory sampling in online learning can result in undesirable outcomes. For example, fluctuating decisions can reduce consumer trust in the decision-making system, decisions that worsen over time (from the perspective of the user) can lead to user abandonment, and decisions which increase disparities in outcomes across groups or individuals can be viewed as unfair. Incorporating trajectory constraints can address these issues, although algorithm design and analysis is non-trivial, even for simple settings. In this talk, lay the groundwork for monotonic stochastic convex optimization and discuss challenges in more general settings.

MB10

Summit - 330

Session on Networks and Platform Design

Invited Session

Auctions and Market Design

Chair: James Siderius, Tuck School of Business at Dartmouth College, Hanover, NH, United States

Co-Chair: Mohamed Mostagir, University of Michigan, Ann Arbor, MI, United States

1 - Traffic Network Games with Risk-averse Players

Emerson Melo, Indiana University Bloomington, Bloomington, IN, United States

Congestion in traffic networks is pervasive, with costs soaring from 15 billion dollars in 1982 to 190 billion in 2021. Heavy and uncertain traffic conditions exacerbate commuting experiences for millions across the US. The Wardrop Equilibrium (WE), a foundational concept for analyzing traffic network games, assumes risk-neutral players, neglecting congestion cost variability. To address this, we propose a model incorporating the stochastic nature of congestion costs, focusing on risk-averse users. We model risk-averse behavior using tail risk measures, capturing the influence of rare, heavy traffic events. Our contributions include establishing conditions for risk-averse WE in various networks, quantifying inefficiencies in risk-averse WE, and extending our model to study spatial oligopoly price competition in networked markets affected by congestion. Our analysis highlights the network topology's role, congestion cost functions, and the characteristics of stochastic congestion shocks, providing insights into optimizing traffic flow in uncertain environments. Finally, we discuss how our framework is suitable for implementing data-driven techniques to recover the parameters describing the cost functions and users' degree of risk aversion.

2 - Algorithmic Demand Allocation to Heterogeneous Sellers in Inventory Competition

Shengyu Cao, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu

In the digital commerce era, platforms such as Amazon and PDD employ sophisticated algorithms of demand allocation to optimize operational efficiency and market dynamics. In this paper, we explore the dynamics between platform rules and seller behavior within online marketplaces, particularly focusing on how inventory-level allocation strategies shape market competition and platform intermediation practices. Through a detailed analysis of two sellers offering identical products on a monopolistic platform, we investigate the effects of different allocation rules on seller order decisions, platform payoff, and overall market equilibrium. Contrary to expectations, findings reveal that inventory-agnostic priority rules can mirror monopolistic decision-making scenarios, suggesting that these allocation methods do not necessarily influence the platform's optimal payoff or affect the equilibrium total order quantity. Moreover, the introduction of randomness in demand allocation intensifies competition and influences replenishment strategies, potentially increasing the platform's payoff and altering market dynamics significantly. The introduction of inventory-aware allocation rules offers a novel approach to demand allocation, promoting competitive restocking strategies. Our study further extends to a multiple-seller framework, analyzing how various allocation rules impact seller behavior and platform economics in a more complex market setting. This comprehensive analysis sheds light on the strategic design of allocation mechanisms, highlighting their significant implications for seller competition and platform intermediation in the digital commerce landscape.

3 - Peer Filtering: Democratic Misinformation Control in Social Networks

Calvin Roth, University of Minnesota, Minneapolis, MN, United States, Ankur Mani, Krishnamurthy Iyer

Online social networking platforms suffer from the problem of misinformation spread that has serious societal implications in areas such as politics and public health, among others. Social networking platforms spend a lot of effort on curtailing misinformation and the techniques are often top down involving content moderation. Research on misinformation in social networks has overlooked the ability of the network as a whole to moderate the content in a more democratic way. We study the power of peers in filtering out false information (peer filtering) in social networks and how the platform can assist in peer filtering without direct content moderation or judging individuals. We present a tractable model of content spread in a social network of Bayesian users who derive utility from sharing a content depending upon its veracity and alignment with their opinions. Based on their posterior beliefs, the users choose whether or not to further share the content. We study the resulting equilibrium in the network and the different types of equilibria that emerge. We show that, under all types of equilibria, true content spreads more in the network than false content, but true content cannot go viral without false content also going viral. Using this model, we describe how key metrics of the network vary with the parameters of the model. These metrics include how much false content is successfully filtered out and how much true content remains as well as the volume and engagement in the network.

MB11

Summit - 331

Machine Learning and Urban Network Analysis

Invited Session

Telecommunications and Network Analytics

Chair: Jin-Zhu Yu, UT Arlington, Arlington, TX, United States

1 - Enhancing the Resilience of Large-Scale Infrastructure Networks Using GNN-Enhanced Deep Reinforcement Learning

Jin-Zhu Yu, UT Arlington, Arlington, TX, United States

Infrastructure systems are vital for socio-economic prosperity and public safety. Optimizing the restoration sequence of damaged components in capacitated infrastructure networks after disruptions is crucial. However, this task is an NP-hard problem that cannot be solved efficiently, particularly for large-scale networks. This study introduces a GNN-enhanced deep reinforcement learning (GNN-DRL) framework to strengthen the resilience of large-scale infrastructure networks. The GNN model effectively learns network embeddings, while the DRL model optimizes repair decisions and network flow. Extensive numerical experiments on various real-world and synthetic large-scale networks under multiple hazard scenarios are conducted to validate the GNN-DRL model. When compared to heuristic approaches like greedy algorithms and genetic algorithms, the GNN-DRL model outperforms them in solution quality, resulting in higher system resilience.

2 - Generating Graph Flows via Diffusion-Based Generative Model

Yu Wang, University of Oregon, Eugene, OR, United States, Jinzhu Yu

Real-world infrastructure systems, such as power grids, water supply networks, telecommunication systems, and transportation routes, are inherently structured as complex graphs. These networks are composed of nodes that represent crucial control points, such as stations or operators, and edges that depict the flow of resources or data. Accurate modeling of these infrastructure graphs is essential for a range of applications, including enhancing infrastructure design, optimizing flow management, and contributing to social science research. A primary challenge in this field is capturing the distinct graph topology and the intricate dynamics dictated by domain-specific physical laws. In response to this challenge, we introduce "Diff-Infra," a cutting-edge diffusion-based generative model designed to create realistic infrastructure graphs. The model operates in two phases: In the forward phase, Diff-Infra systematically introduces noise to degrade the graph's structure, while in the reverse phase, a graph neural network is trained to remove this noise, effectively learning to reconstruct and denoise the infrastructure graphs. Additionally, physical constraints are rigorously integrated to ensure the model adheres to realistic dynamics. Our experimental findings confirm that Diff-Infra adeptly replicates both the local and global characteristics of infrastructure graphs. It accurately models essential features such as flow conservation, cyclic trends, and congestion at critical junctions, demonstrating its potential to significantly improve infrastructure planning and operations.

3 - Hazard, Infrastructure, and Community Resilience Network Models

Kenneth Harrison, National Institute of Standards and Technology (NIST), Gaithersburg, MD, United States, Tasnim Faiz, William Hughes, Zeinab Farahmandfar

Natural disasters, like floods, tornadoes, earthquakes, and wildfires, damage a community's critical infrastructure and buildings and disrupt its daily functions. Communities must prioritize resource allocation to take mitigation actions to reduce the impact of disasters while constrained by budgetary and other restrictions. For effective community resilience planning, a modeling framework incorporating structural engineering and the physical and social sciences is necessary to quantify the impact of uncertain hazards on the built environment and translate these impacts into societal outcomes. This work presents a novel decision-making framework that leverages existing probabilistic relationships throughout for simulating hazards, modeling the response of engineered structures, assessing impacts on building functionality, and ultimately, estimating whether societally meaningful objectives will be met, such as surviving the hazard event and remaining in one's home. The optimization model underlying the framework utilizes the network structures of the interdependent infrastructure and social systems to aid the decision-making process for critical infrastructure and buildings, considering varied scenarios and restrictions. Case studies of different communities facing various hazards are presented to show the adaptability of the model and decision framework.

4 - Machine Learning for Disrupting Max Flow Networks

N. Orkun Baycik, University of Denver, Denver, CO, United States

This study presents machine learning based approaches (MLBA) to solve the maximum flow network interdiction problem (MFNIP). The mathematical formulation of this NP-complete problem is an Integer Program (IP) and large size instances can be computationally expensive to solve. As an alternative, this study applies tree-based learning methods (the decision tree and random forest regression techniques) to various sizes of networks to analyze the solution times as well as the optimality gaps compared to the IP formulation of MFNIP. The proposed methods rely on predicting disruptions that individual arc interdictions can cause on the network. Without solving an optimization model, the decision-maker can choose a set of arcs with high predicted disruptions to cause the maximum damage. They can also produce insights into the most vulnerable components of a network using this approach. The results indicate that MLBA are able to find solutions within seconds and increasing the network size does not affect these solution times. The quality of the solutions are acceptable for the majority of the instances especially for small to medium interdiction budget levels. For large-size instances, the results indicate that MLBA are efficient and effective, whereas solving an IP model can require long solution times and consume a significant amount of computer memory.

5 - Enhancing Decision-making and Representation of Complex Infrastructure Systems Using Advanced Graph Neural Networks

Xudong Fan, SUNY at Buffalo, Buffalo, NY, United States, Jürgen Hackl

Aging urban infrastructure systems pose risks to urban daily functions, resulting in economic losses and public health concerns when failures occur. To enhance the sustainability and resilience of urban infrastructure systems, public utilities are actively seeking optimal strategies for infrastructure management. However, making optimal decisions for a resilient infrastructure system is difficult due to the dynamic behaviors. In this study, we first introduced the Deep Reinforcement Learning (DRL) and Graph Convolutional Networks (GCN) for the resilience-oriented decision-making in urban infrastructure systems. We will then introduce our studies about developing efficient graph learning algorithm for accurate network representation of complex infrastructure systems, namely regional spatial graph convolutional network (RSGCN). The performance of developed RSGCN model is evaluated by modeling three spatially embedded network datasets, including a synthetically made network dataset and two real-world infrastructure network datasets.

6 - Strategic Cyber Defense: Game-Theoretic Penetration Testing, Risk Assessment, and Resilience in 5G-Enabled Networks

Quanyan Zhu, New York University, New York, NY, United States

Penetration testing assesses an organization's IT infrastructure's security posture and its resilience against cyber threats. It helps strategically design cyber defense to counter vulnerabilities and bolster security. This strategic defense can be modeled using game-theoretic and learning-based methods. This approach facilitates the development of agent-based technologies for automating strategic penetration testing and proactive defenses. This talk aims to address challenges in modeling, analyzing, and computing dynamic game models in modern industrial control systems integrated with 5G networks and cloud services. We propose new game-theoretic solution concepts to capture the networks' dynamic and multiscale attributes. The talk will discuss new learning-based and computationally scalable algorithms to advance penetration testing technologies. Integrating control, learning, and game-theoretic methodologies provides a quantitative approach to assess risks, enhance resilience, and guide policy-making, including standards and cyber insurance.

MB12

Summit - 332

Agricultural Economics and Food Security Analytics

Flash Session

Chair: Hengyu Liu, Beijing University of Posts and Telecommunications, Beijing, N/A

1 - Price Subsidies with OR Without Physical Procurement: Impact on Quality, Profits, and Welfare

Omkar D. Palsule Desai, Indian Institute of Management Indore, Indore, India, Aysajan Eziz, Srinagesh Gavirneni

Newly introduced price subsidy programs without physical procurement have resulted in an increased (vis-\`{a}-vis the subsidy programs with physical procurement) preponderance of deliberate quality degradation for certain crops by Indian farmers. The physical procurement of crops from farmers provides an alternate sales channel for farmers, enabling the government to subdue the competition between farmers in the open market. The farmers' deliberate quality degradation may moderate the benefits of the altered competitive structure. The viability of such programs has been questioned, and policymakers are looking for guidance. Using a multi-stage incomplete information-based (Bayesian) game-theoretic model, we comprehensively characterize the farmers' strategic production and selling decisions. We demonstrate that the government's price support with multiple sales channels created by physically procuring the crop improves the producer surplus when the farmers' landholdings are highly disparate and the quality-based competition is lower. A higher minimum support price (MSP), a higher landholding for a farmer producing a high-quality crop, and a lower high-quality crop price premium in the open market improve the consumer surplus, the social surplus, and the average quality of the crops supplied by farmers when the government does not procure the crop physically v/s when it does. To eliminate farmers' deliberate crop quality degradation, when MSP is moderate, if not lower, the government should (should not) procure the crop physically if the high-quality crop price premium in the open market is higher (lower). We provide guidance to the government exercising caution in selecting MSP and the farmer subsidy program.

2 - Stochastic Dominance in Cultivar Selection: Addressing Climate Change Uncertainties for Resilient Agriculture

Shayan Tohidi, Iowa State University, Ames, IA, United States, Sigurdur Olafsson

Food supplies are at risk due to climate change, which can affect their accessibility and quality. Extreme and unpredictable weather events recognize climate change while their patterns also change, implying more uncertainty that leads to higher risk. This calls for better methods for cultivar selection that can both consider risk due to current variability of environments, and the increased risk due to changes in environments caused by climate change. We refer to probabilistic ranking as ordering cultivars based on their probability to perform better across a set of environments versus their mean performance. Such probabilistic rankings address both mean performance (e.g., mean yield) and uncertainties due to environments and potentially also environmental changes. Motivated by the need for better methods to select crop cultivars that are resilient to change in environments caused by climate change, we have proposed a novel probabilistic ranking method based on stochastic dominance, a partial ranking between two random variables. By exploring all cultivar pairs, this method can order cultivars by their yield distribution in a computationally efficient manner. For each, we identify the superior one for a risk-averse decision-maker. It also helps us obtain a sensitivity analysis for each pair of cultivars, so we know beforehand that changing the distributions will determine how the result will be varied. Thus, by estimating current and future climate accurately, we can find the best cultivars for now and the appropriate potential adjustments for the future, leading to a more thoughtful decision in response to climate change.

3 - Early Harvest Behavior in Fruit Supply Chains

Hengyu Liu, Beijing University of Posts and Telecommunications, Beijing, China, People's Republic of

In developing countries, fruit farmers are price takers and the market price is determined by the aggregate supply. To reduce price (or inventory) pressure, many farmers harvest their fruit early, when it is not yet fully ripe. Also, farmers face two types of uncertainties: (*i*) yield and (*ii*) unripe quality uncertainty in the early harvest season. Usually, the farmers' early harvest behavior (EHB) will lead to the effect of negative word of mouth (WOM), decreasing the subsequent demand for ripe fruit. By accounting for the competition among farmers, yield and unripe quality uncertainties, and WOM, we develop a game-theoretical model to examine the fruit supply chain and market dynamics that engender EHB as well as to analyze its implications to different stakeholders. We fully characterize the farmers' equilibrium planting decisions and EHB. We demonstrate that the EHB can be eliminated only when the realized yield is moderate and is of relatively low quality. Contrary to the prevailing intuition that EHB hurts consumers because they are enticed to consume unripe (i.e., low-quality) fruit, we show that this behavior can increase their welfare under certain conditions due to the higher supply. Furthermore, we caution that the famers' EHB can make themselves worse off if they face a moderate degree of yield uncertainty. In general, if there are a small number of famers competing in the market while facing a high degree of yield uncertainty, EHB can make the farmers and consumers achieve a win-win situation in the market.

MB13

Summit - 333

Games and Markov Systems

Flash Session

Chair: Ozgun Caliskan Demirag, Penn State Erie, The Behrend College, Erie, PA, United States

1 - Dynamic Programming for Efficient Semiparametric Estimation: Control of Universal Least Favorable Flows

KAIWEN HOU, Columbia Business School, New York, NY, United States

To aid reliable and efficient causal estimation, previous literature has proposed constructing normalizing flows in a statistical manifold, with tangent vectors aligned with the efficient influence function of the estimand. This paper formulates the same problem from the dynamic programming perspective. Starting from an initial suboptimal estimator of the probability measure (which may be highly biased), the next optimal perturbation step is determined by locally aligning with the efficient influence function evaluated at that point.

By adopting a dynamic programming approach, we extend this perturbation to a global scale, enabling the identification of an integral curve on the statistical manifold that passes through the initial estimator. This integral curve essentially defines a universal least favorable flow, corresponding to the tightest Cramér-Rao lower bound for the estimation problem. Dynamic programming techniques allow us to control this flow effectively, ensuring each step minimizes estimation variance and enhances efficiency.

Furthermore, theoretical guarantees are provided, demonstrating the robustness and optimality of our approach. Specifically, we establish a local perturbation error bound, ensuring that each step's deviation from the true efficient influence function is controlled. Additionally, a global approximation error bound is derived, indicating the cumulative error over the entire estimation process. We also establish the consistency and asymptotic normality of the estimators obtained through our method, showing that they achieve the semiparametric efficiency bound. Preliminary simulation studies related to the estimation of the population mean confirm the effectiveness of our approach, consistently yielding lower mean-squared errors compared to traditional methods such as targeted learning.

2 - Near-Optimal Cost Function Approximation for Multi-Period Technician Routing and Scheduling

Kaiwen Li, Beijing Foreign Studies University, Beijing, China, People's Republic of, Xi Chen

We study the problem of designing near-optimal formulation of cost function approximation (CFA) for multi-period technician routing and scheduling problem in an online learning setting. Prior works using CFA design the parameterized model based on domain knowledge and tune parameters in an offline setting without providing error analysis. However, when the impact of uncertainty is easy to recognize, we propose an approach to find a near-optimal formulation of CFA a prior. More precisely, we first build Kullback-Leibler ambiguity set of future cost incurred by certain action, then we use the relationship between regularization and robust optimization to construct an equivalent regularization term, which leads to the modified cost function in CFA. Next, the performance gap between our proposed approach and optimal policy is given by error analysis of state aggregation. Numerical experiment demonstrates that the proposed approach (without tuning parameters) performs close to optimal in various settings of urban delivery.

3 - Optimal Co-production for Data-Intensive Goods

Bozhuang Lei, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Gang Li, Yimin Yu

Many firms owning reams of production or consumer usage data lack advanced AI capabilities and thus seek partnerships with AI-powered platforms to co-produce data-intensive products such as automotive vehicles. Using a game-theoretical model, we characterize the optimal data-algorithm co-production scheme between the firm and platform, deriving the optimal data size and algorithm construction level respectively contributed by the firm and platform. We also specify the output requirement and transfer payment under the co-production. In the single-period setting, our analysis shows that the firm is willing to build the partnership when the data leakage cost is low. In the two-period setup with data collection, we find that the firm strategically lowers its first-period price to gather more consumer data and can adopt either fully sharing or withholding strategy for data sharing in the second period. The firm prefers cooperation only when the data leakage cost is low and the data collection efficiency is high. We also show that the firm benefits more from data-hungry or data-efficient algorithms than those with a relatively balanced data and algorithm importance. Overall, this study investigates the market implications of co-production for AI-powered products and highlights the roles of data-algorithm interaction features in driving the firm profitability.

4 - Effects of cost information asymmetry on business model choice of online platforms

Bingru Wang, Xi'an Jiaotong University, Xi'an Shaanxi, China, People's Republic of, Jun Lin

We utilize a game-theoretic model to examines how the strategic interaction between a manufacturer and an online platform determines the channel structure and how private cost information affects the platform strategy decision. First, we study the platform model selection strategies under symmetry cost information. We find that in most cases, the resale model remains the optimal choice for platforms, unless the agency model features higher platform fees and lower manufacturer costs. And there always exists a certain range where both platform and manufacturer prefer either the reselling or agency model, achieving a win-win situation. Second, we utilize signaling game analysis to examine the impact of cost information asymmetry on platform strategic decisions and supply chains. We find that type-L manufacturers, in order to distinguish themselves from other types, tend to actively distort prices downward, deviating from the optimal price. This results in a decrease in profits and efficiency for supply chain members under the agency model, while consumer surplus increases. Unlike symmetric cost information, there exists only a range where both parties prefer the reselling model under cost information asymmetry. Finally, we discuss the platform strategy shifts under cost information asymmetry and find that with the increase in commission rate of agency model (or proportion of type-L manufacturers), the range of platform preference for reselling model expands.

5 - Innovation and pricing competition with absorptive capacity

Xiuli He, UNC Charlotte, Charlotte, NC, United States, Huimin Zhan, Xianjin Du, Hong Fu

We develop a game theoretical model of sequential entrance in which an entrant with absorptive capacity tries to enter a market where the innovator exists. The innovator decides on innovation and pricing strategies to prevent or allow the market entry of the absorptive entrant. We show that the entrant's absorptive capacity negatively affects the innovator's pricing strategy, which may even sell at a lower price than that of the entrant. However, the innovation efforts may increase to establish higher barriers to invasion. The innovator may have to allow the market entry if the entrant has a stronger absorptive capacity or the competition between products is lower. Interestingly, the entrant has a second-mover advantage under certain conditions. Our findings reveal the effect of absorptive capacity and competition in operational decisions and provide practical recommendations for decision-making by competing manufacturers.

6 - Contracts for Procurement under Spot Price Fluctuations

Ozgun Caliskan Demirag, Penn State Erie, The Behrend College, Erie, PA, United States, Yong-Hong Kuo, Jie Wang

Employing a three-stage game-theoretic model, we delineate the equilibrium strategies of suppliers and manufacturers engaged in procurement option contracts amid price volatility and breach contingencies. We find that contract effectiveness diminishes with demand and spot price variability but rises with the supplier's spot market search costs. Notably, flexibility in renegotiation outperforms models presuming no breaches. Our findings offer crucial insights for businesses reliant on product trading. Despite breach risks, well-designed option contracts tailored to address price fluctuations can enhance supply chain profitability. Notably, renegotiation schemes yield maximal benefits.

MB14

Summit - 334

Tensor Factor Models in Finance and Business

Invited Session

Finance

Chair: Elynn Chen, New York University, New York, NY, 10012, United States

1 - 3D-Pca: Factor Models with Restrictions

Martin Lettau, UC Berkeley, Berkeley, CA, United States

This paper proposes latent factor models for multidimensional panels called 3D-PCA. Factor weights are constructed from a small set of dimension-specific building blocks, which give rise to proportionality restrictions of factor weights. While the set of feasible factors is restricted, factors with long/short structures often found in pricing factors are admissible. I estimate the model using a 3-dimensional data set of double-sorted portfolios of 11 characteristics. Factors estimated by 3D-PCA have higher Sharpe ratios and smaller cross-sectional pricing errors than models with PCA or Fama-French factors. Since factor weights are subject to restrictions, the number of free parameters is small. Consequently, the model produces robust results in short time series and performs well in recursive out-of-sample estimations.

2 - Shrinking the Term Structure

Markus Pelger, Stanford University, Stanford, CA, United States, Damir Filipovic, Ye Ye

We propose a new framework to explain the factor structure in the full cross section of Treasury bond returns. Our method unifies non-parametric curve estimation with cross-sectional factor modeling. We identify smoothness as a fundamental principle of the term structure of returns. Our approach implies investable factors, which correspond to the optimal spanning basis functions in decreasing order of smoothness. Our factors explain the slope and curvature shapes frequently encountered in PCA. In a comprehensive empirical study, we show that the first four factors explain the time-series variation and risk premia of the term structure of excess returns. Cash flows are covariances as the exposure of bonds to factors is fully explained by cash flow information. We identify a state-dependent complexity premium. The fourth factor, which captures complex shapes of the term structure premium, substantially reduces pricing errors and pays off during recessions.

3 - Stochastic Low-Rank Tensor Bandits for Multi-Dimensional Online Decision Making

Emma Jingfei Zhang, Emory University, Atlanta, GA, United States

Multi-dimensional online decision making plays a crucial role in many real applications such as online recommendation and digital marketing. In these problems, a decision at each time is a combination of choices from different types of entities. To solve it, we introduce stochastic low-rank tensor bandits, a class of bandits whose mean rewards can be represented as a low-rank tensor. We consider two settings, tensor bandits without context and tensor bandits with context. In the first setting, the platform aims to find the optimal decision with the highest expected reward, a.k.a, the largest entry of true reward tensor. In the second setting, some modes of the tensor are contexts and the rest modes are decisions, and the goal is to find the optimal decision given the contextual information. We propose two learning algorithms tensor elimination and tensor epoch-greedy for tensor bandits without context, and derive finite-time regret bounds for them. Comparing with existing competitive methods, tensor elimination has the best overall regret bound and tensor epoch-greedy has a sharper dependency on dimensions of the reward tensor. Furthermore, we develop a practically effective Bayesian algorithm called tensor ensemble sampling for tensor bandits with context. Extensive simulations and real analysis in online advertising data back up our theoretical findings and show that our algorithms outperform various state-of-the-art approaches that ignore the tensor low-rank structure.

4 - Distributional Matrix Completion via Nearest Neighbors in the Wasserstein Space

Jacob Feitelberg, Columbia University, New York, NY, United States

We introduce the problem of distributional matrix completion: Given a sparsely observed matrix of empirical distributions, we seek to impute the true distributions associated with both observed and unobserved matrix entries. This is a generalization of traditional matrix completion where the observations per matrix entry are scalar valued. To do so, we utilize tools from optimal transport to generalize the nearest neighbors method to the distributional setting. Under a suitable latent factor model on probability distributions, we establish that our method recovers the distributions in the Wasserstein norm. We demonstrate through simulations that our method is able to (i) provide better distributional estimates for an entry compared to using observed samples for that entry alone, (ii) yield accurate estimates of distributional quantities such as standard deviation and value-at-risk, and (iii) inherently support heteroscedastic noise. We also prove novel asymptotic results for Wasserstein barycenters over one-dimensional distributions.

MB15

Summit - 335

Matchings in Random Graphs

Invited Session

Revenue Management and Pricing

Chair: Will Ma, Columbia Business School, New York, United States

Co-Chair: Pranav Nuti, University of Chicago, Chicago, IL, United States

Co-Chair: Calum MacRury, Columbia University, New York, NY, United States

1 - Query Efficient Weighted Stochastic Matching

Mohammad Saneian, Northeastern University, Brookline, MA, United States, Mahsa Derakhshan

In this talk, we discuss the *weighted stochastic matching* problem. Let $G=(V, E)$ be a given edge-weighted graph and let its realization $\mathcal{G}$ be a random subgraph of G that includes each edge $e\in E$ independently with a known probability p_e. The goal in this problem is to pick a sparse subgraph Q of G without prior knowledge of G's realization, such that the maximum weight matching among the realized edges of Q (i.e., the subgraph $Q \cap \mathcal{G}$) in expectation approximates the maximum weight matching of the entire realization $\mathcal{G}$.

In this work, we present a 0.68 approximation algorithm with $O(1/p)$ queries per vertex, which is asymptotically tight. This is even an improvement over the best-known approximation ratio of $2/3$ for unweighted graphs within the $\text{poly}(1/p)$ regime due to Assadi and Bernstein [SOSA'19]. The $2/3$ approximation ratio is proven tight in the presence of a few correlated edges in $\mathcal{G}$, indicating that surpassing the $2/3$ barrier should rely on the independent realization of edges. Our analysis involves reducing the problem to designing a randomized matching algorithm on a given stochastic graph with some *variance-bounding* properties.

2 - Fundamental Limits for Broken Sample Problems

Jiaming Xu, Duke University, Durham, NC, United States

In this talk, I will explore a classic statistical challenge: the broken sample problem. Imagine a random sample of vectors is drawn from a joint distribution. Before observing them, each vector in the sample is broken into two subvectors and the original pairing information is lost. The central question we address is whether it is possible to recover this lost pairing information just by observing these two sets of subvectors. The broken sample problem has a rich history in statistics, dating back to the 1960s, with initial applications in matching records for surveys and censuses. Recently, this problem has gained renewed interest due to its relevance in fields such as data de-anonymization, robotics, multi-target tracking, signal processing, and data integration. It encapsulates a range of related problems found in the literature, including record linkage, database alignment, geometric matching, and shuffled regression. Despite extensive research over the past decade, determining the precise statistical thresholds for detection and estimation has remained a challenge. In this presentation, I will discuss our latest progress in determining these thresholds and important open problems on this topic.

3 - Online Matching in Geometric Random Graphs

Flore Sentenac, HEC Paris, Jouy en Josas, France, Vianney Perchet, Nathan Noiry, Laurent Ménard, Matthieu Lerasle

In online advertisement, ad campaigns are sequentially displayed to users. Both users and campaigns have inherent features, and the former is eligible to the latter if they are ``similar enough''. We model these interactions as a bipartite geometric random graph: the features of the 2N vertices (N users and N campaigns) are drawn independently in a metric space and an edge is present between a campaign and a user node if the distance between their features is smaller than c/N, where c>0 is the parameter of the model. Our contributions are two-fold. In the one-dimensional case, with uniform distribution over the segment [0,1], we derive the size of the optimal offline matching in these bipartite random geometric graphs, and we build an algorithm achieving it (as a benchmark), and analyze precisely its performance. We then turn to the online setting where one side of the graph is known at the beginning while the other part is revealed sequentially. We study the number of matches of the online algorithm closest, which matches any incoming point to its closest available neighbor. We show that its performances can be compared to its fluid limit, completely described as the solution of an explicit PDE. From the latter, we can compute the competitive ratio of closest.

4 - Online Matching for Edge Arrivals with Vanishing Probabilities

Pranav Nuti, Stanford University, Stanford, CA, United States, Will Ma, Calum MacRury

Temporary abstract.

MB16

Summit - 336

New Approaches for Matching Markets

Invited Session

Revenue Management and Pricing

Chair: Francisco Castro, UCLA Anderson School of Management, Los Angeles, CA, United States

Co-Chair: Alfredo Torrico, Cornell University, Ithaca, NY, United States

1 - Potential-Based Greedy Matching for Dynamic Delivery Pooling

Matias Romero, Columbia University, New York, NY, United States, Hongyao Ma, Will Ma

We study a dynamic non-bipartite online matching problem motivated by applications in mobility platforms. Food delivery platforms, in particular, implement real-time order pooling strategies to optimize operational efficiency and sustainability. These decisions involve a trade-off between immediate reward and potentially better future pooling opportunities. Existing heuristics often stem from general online matching algorithms but vary in effectiveness across applications. This paper introduces a new potential-based greedy (PB) algorithm, which balances instant reward with the potential loss incurred by matching non-ideal job pairs, without relying on forecasts or partial information about future arrivals.

We analyze the theoretical performance of PB in a dynamic non-bipartite matching model under different reward structures. Our findings demonstrate that PB significantly improves upon the naive greedy approach in terms of worst-case performance, particularly for reward functions that capture the distance saved by pooling deliveries. Empirical validation through simulations and order-level data from a food delivery platform shows that PB outperforms commonly used heuristics. These results highlight PB's robustness and practical effectiveness, offering a promising solution for real-time decision-making in diverse dynamic matching scenarios.

2 - Dynamic Matching with Post-Allocation Service and Its Application to Refugee Resettlement

Soonbong Lee, Yale University, New Haven, CT, United States, Kirk Bansak, Vahideh Manshadi, Rad Niazadeh, Elisabeth Paulson

Motivated by our collaboration with a major refugee resettlement agency in the U.S., we study a dynamic matching problem where each new arrival (a refugee case) must be matched immediately and irrevocably to one of the static resources (a location with a fixed annual quota). In addition to consuming the static resource, each case requires post-allocation services from a server, such as a translator. Given the uncertainty in service time, a server may not be available at a given time, thus referred to as a dynamic resource. Upon matching, the case waits to avail service in a first-come-first-serve manner. Bursty matching may result in undesirable congestion of the servers. Consequently, the planner (the agency) faces a dynamic matching problem with an objective combining the matching reward (pair-specific employment outcomes) with the cost for congestion for dynamic resources and over-allocation for the static ones. Motivated by the observed fluctuations in the refugee pools across the years, we aim to design algorithms with no distributional knowledge. We develop learning-based algorithms that are asymptotically optimal in certain regimes, easy to interpret, and computationally fast. Our design is based on learning the dual variables of the underlying optimization problem; however, the main challenge is the time-varying nature of the dual variables associated with dynamic resources. Our theoretical development integrates techniques from Lyapunov analysis, adversarial online learning, and stochastic optimization. When tested on our partner agency's data, our method outperforms existing ones making it a viable candidate for replacing the current practice upon experimentation.

3 - Constrained Contracts in Sequential Duopolies

Woo-Jin Kim, University of Southern California, Los Angeles, CA, United States, Afshin Nikzad, Guofu Tan

We study a model of duopolistic competition in which two sellers sequentially offer sales contracts with one unit of the divisible good to a buyer with private information. The first-mover can choose an arbitrary contract (i.e., a potentially uncountable set of price-quantity pairs) whereas the contract-constrained follower can only make a take-it or leave-it offer (i.e., selling one unit of the good for a fixed price). We characterize the equilibrium of the game and find that (i) the optimal mechanism for the first-mover also involves a take-it or leave-it offer but with a smaller quantity and a lower price than the second-mover, and (ii) the follower earns higher revenue than that of the first-mover, despite the constrained contract space it can choose from. Moreover, (iii) we show that the presence of the second seller reduces the first (incumbent) seller's profit and increases the buyer's surplus as well as the total surplus.

4 - Algorithmic Solutions for Choice-Based Matching Markets

Alfredo Torrico, Cornell University, New York, NY, United States

Choice-based matching platforms have recently proliferated thanks to their application in outsourcing, dating, accommodation and carpooling. These markets are generally characterized by: (i) two opposite sides who are looking to match and (ii) their decentralized nature in which users can freely choose someone (or not) from the displayed alternatives. Platforms, then, serve as a mediator that facilitates the matching process between both sides by controlling what users see and how they interact. One of the main challenges that these platforms face is choice congestion: If the display of alternatives is not carefully designed, then the most popular users may get more requests than they can actually serve. Two-sided assortment optimization has recently gained significant attention as an algorithmic tool to reduce congestion. An ideal algorithm balances between offering enough relevant alternatives to keep users engage and controlling the display of popular options to minimize congestion. In this talk, I will present recent advances in two-sided assortment optimization such as how different platform designs impact the market's outcome and the algorithmic challenges that each design poses.

MB17

Summit - 337

Consumer Choice and Search

Invited Session

Revenue Management and Pricing

Chair: Aydin Alptekinoglu, The Pennsylvania State University, University Park, PA, 16801, United States

1 - Avoiding Unrealistic Behavior in ScanPro-like Demand Models

Arian Aflaki, University of Pittsburgh, Pittsburgh, PA, United States, Deniz Eskandani, Su-Ming Wu

Empirical multiproduct demand models such as ScanPro can exhibit undesirable properties due to their log-linear nature. For example, increasing the price of all products can increase total demand. We propose an approach combining choice and empirical models to avoid these undesirable properties while maintaining ease of estimation.

2 - Dynamic Assortment in Make-To-Order Systems

Yuan Guo, George Washington University, DC, DC, United States, Chen-An Lin

This paper studies a dynamic assortment problem under a make-to-order system. The firm determines an assortment to each incoming customer, where customers arrive following a Poisson process and decide whether to purchase or balk upon seeing the assortment and the system congestion level. Products vary in production time, cost, and selling price. The make-to-order system operates as a first-come-first-served single-server queue. If a customer chooses to buy, they join the queue until production is complete while their order is added to the system congestion level as the updated system state. The firm aims to maximize long-run average profit rates by selecting assortments for different congestion levels. We show that, under linear product profit in production time, the optimal assortment policy aligns with revenue-ordered sets. We find that, typically the optimal assortment increases with congestion levels, but the full structure suggests it can first expand and then contract, indicating a delicate balance between congestion and profit maximization.

3 - Rational Choice Models: The Temporal Tree Representation

QI FENG, Purdue University, WEST LAFAYETTE, IN, United States, J Shanthikumar, MENGYING XUE

Choice models, specifying the consumer's choice probability of an option over a given choice set, are widely studied and applied in many fields. We propose a temporal tree representation of choice that covers all rational choice models. Compared with the existing structural choice models, the tree representation exhibits two major advantages that overcome the major challenges of model identification. First, all rational choice models have a tree representation, and a subclass of tree representation (with set-dependent branching) has a one-to-one correspondence to the rational choice models. This bridges the gap in the existing characterizations of structural models, which are unidentifiable, impose uninterpretable conditions, or do not cover the entire space of rational choice models. Second, the tree representation allows for the flexibility of systematically specifying the choice model structure based on available knowledge and data. In particular, the number of parameters needed to specify a tree representation can be primarily determined by the sufficient knowledge level, which corresponds to a specific layer of the tree branching. The sufficient knowledge level can be empirically determined based on the amount of available data, which in turn determines the number of parameters needed for model estimation. Therefore, the tree representation allows for a natural way of data integration, avoiding misspecification due to restrictive assumptions and overfitting for general models.

4 - Discrete Choice via Sequential Search

Aydin Alptekinoglu, The Pennsylvania State University, University Park, PA, United States

We study the information collection or search phase and its impact on final choices in a retail shopping context. To this end, we embed an analytically tractable discrete choice model in a classical model of sequential search with perfect recall. Although significant progress has been made in the theory literature in analyzing consumers' discrete choice behavior using random utility models, discrete choice through a sequential search process has not received enough attention due to analytical intractability issues. We build on the seminal *Pandora's Problem* introduced by Weitzman (1979) as a model of sequential search with perfect recall, and on the *Exponomial Choice* model (Alptekinoglu and Semple 2016) as a model of discrete choice (with each choice specified by a deterministic or observable utility component, minus a random utility component following an exponential distribution). We derive the search path and final choice probabilities in closed form, develop all the analytical tools to optimize prices for a given assortment of products, and show that the optimal assortment must contain some number of highest-value products at optimal prices. These results enable joint optimization of assortment and prices efficiently in retail shopping scenarios that can be well represented by sequential search. The structure of the solution accommodates a distinct group of products that are priced just so they remain on the search path with higher probability. We also conduct a simulation study to compare our model with the state-of-the-art in empirical modeling of sequential search.

MB18

Summit - 338

The Theory and Practice of Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Pin Gao, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Xiaotang Yang, University of Toronto, Toronro, ON, M5G 1N6, Canada

1 - Online Learning for Pricing in On-Demand Vehicle Sharing Networks

Huanan Zhang, University of Colorado Boulder, Boulder, CO, United States, Saif Benjaafar, Xiangyu Gao, Xiaobing Shen

We consider the pricing problem in on-demand vehicle sharing networks with online demand learning. When there is no prior information available on the demand functions, the main challenge in designing an online learning algorithm is how to explore the demand functions while maintaining a balanced network. We address this challenge with an online learning algorithm adapted from the ellipsoid method. In our algorithm, the search subroutine is based on the idea of bisection and the Upper Confidence Bound, which can locate the price associated with a desired demand level for each type of trip, as characterized by the trip origin and destination, and estimate the gradient information at the price point. By carefully selecting the center of the ellipsoid for each iteration, we can ensure that the expected revenue improves and maintain a balanced network in each iteration. We prove that the regret of our learning algorithm is bounded by \tilde O(sqrt{T}) given a fixed workload parameter Delta. The numerical performance of the algorithm is illustrated using synthetic data. We also discuss extensions to the online learning algorithm in which the workload parameter Delta is unknown.

2 - Assortment Optimization for Office Meal Delivery Platforms

Saman Lagzi, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States, Ningyuan Chen, Pin Gao, Sheng Liu, Chenhao Wang

We study the problem that an office meal delivery platform faces every day. Such platforms connect client office workers to restaurants by offering a menu of compatible restaurants to all the workers in an office. The platform's revenue is a fixed percentage of the value of the restaurant orders. The platform incurs the delivery cost. Each restaurant allocates a fixed capacity to the platform. We model the problem as a join assortment optimization. We prove the problem is APX-Hard while its special cases with only one restaurant or one client are NP-hard.

Using a Linear Programming relaxation of the original model, and by leveraging the rather soft nature of the capacity constraints in practice, we devise an asymptotically optimal randomized rounding algorithm that may breach any capacity constraint only by a negligible probability, as the number of clients and the restaurant capacities grow.

In collaboration with Canada's largest office meal delivery platform, we test the performance of our algorithm in a controlled switchback experiment in the city of Vancouver. We compare the performance of our methodology with that of the platform, before and after the experimental period. Our results suggests that our methodology improves platform's profit by over 9%, and it improves platform's revenue by nearly 7%.

3 - Benefit of Opaque Selling for Inventory Management

Mingyang Fu, National University of Singapore, Singapore, Singapore, Xiaobo Li

Opaque selling is a sales strategy where the seller does not disclose some specific details of a product or service until after a purchase is made. This approach can effectively generate sales and reduce surplus inventory. In this paper, we consider the case of balancing holding and backlog cost. We begin by examining the inventory costs under different opaque selling schemes and allocation policies. Our findings reveal that the benefits of opaque selling in inventory management is attributed from both allocation flexibility and the demand pooling effect. We show that even a modest degree of flexibility in general opaque selling can result in nearly constant relative inventory cost savings as market size expands. In terms of the profit, we show that when customers' utilities of products adhere to Salop's circle model or are independently and identically distributed, the mixed opaque selling with allocation flexibility consistently yields higher profits by optimizing the balance between the inventory costs saving and the potential revenue loss.

4 - Distributionally Robust Newsvendor on a Metric

Ayoub Foussoul, Columbia University, New York, NY, United States, Vineet Goyal

An e-commerce seller operates multiple warehouses and faces sequentially arriving demand from different locations on a metric. The demand is uncertain with only the first and second moments available. The goal of the seller is to exploit the benefits of risk pooling over demand uncertainty between the warehouses and jointly decide on an initial inventory allocation and demand fulfilment policy that minimize the total expected inventory and fulfillment costs under the worst-case demand distribution. This is a fundamental generalization of the classical distributionally robust newsvendor problem of Scarf (1958). We design a near-optimal policy for the problem with theoretical guarantees on its performance. Our policy generalizes the solution of Scarf (1958), maintaining its simplicity and interpretability: it identifies a hierarchical set of clusters, assigns a “virtual” underage cost to each cluster, then makes sure that each cluster holds at least the inventory suggested by Scarf’s solution if the cluster behaved as a single point with “virtual” underage cost and original overage cost. As demand arrives sequentially, our policy fulfills orders from nearby clusters, minimizing transshipment costs, while balancing inventory consumption across the clusters to avoid depleting any single one. In addition to its theoretical performance, numerical experiments show that our policy performs well in practice.

MB19

Summit - 339

Advances in Market Algorithms and Design

Invited Session

Revenue Management and Pricing

Chair: Anil Omer Saritac, Singapore Management University, Singapore, NW16BJ, Singapore

1 - Multi-Homing Across Platforms: Friend or Foe

Tolga Dizdarer, Carroll School of Management, Boston College, Boston, MA, United States, Gerard Cachon, Gerry Tsoukalas

The gig economy offers flexible work opportunities where contractors enjoy the freedom to control when and how they do their work. In this project, we look at one particular form of freedom: ability to work on multiple platforms, a practice termed multi-homing. We study whether it is in the best interest of the platforms to allow multi-homing and what users have to gain from it, if any. The answer depends on the characteristics of the market. With multi-homing, a platform has less incentive to invest in supply capacity. It also has less incentive to build scale, dampening the excessive competition between platforms. In markets where the latter effect is prominent, platforms have their cake and eat it too. By allowing multi-homing, platforms extract more revenue from the market and also give their workers more freedom, a key move that strengthens platforms' case against regulators that its workers are free from platform's control. Authors: Gerard Cachon, Tolga Dizdarer, Gerrry Tsoukalas

2 - Attribute-Based Pricing: A Novel Formulation and Convergent Algorithms

Mengzhenyu Zhang, University College London, London, United Kingdom, Christopher Ryan

Attribute-based pricing—giving a price to potential product attributes individually and allowing customers to choose the attributes that form the final product—has been shown to improve customer satisfaction in the hospitality industry. In this paper, we consider the problem of finding optimal attribute prices to maximize the expected revenue from selling to a customer who chooses one product from a set of products that differ by only a few attributes. Because of complicated substitution effects among the final products that share common attributes, expected revenue is not concave in attribute prices. Nonetheless, we provide an algorithm to solve the attribute pricing problem and show that it converges to a stationary point that provides a high-quality solution to the problem. Through numerical experiments, we show our algorithm is, on average, ten times faster than gradient-based methods, both in terms of runtime and number of iterations. We also extend our algorithm to a setting where attribute prices are constrained through linear inequalities and prove convergence to a stationary point. We implement our algorithm on a real hotel data set and demonstrate its revenue benefits.

3 - Traffic Management in Live Stream Commerce

Feifan Zhang, Duke University, Durham, NC, United States, Wenchang Zhang, Liu Ming

We study optimizing real-time traffic acquisition for maximizing conversion and profit in live streaming commerce. We introduce a control model emphasizing the strategic use of cash incentives by streamers to attract viewers and enhance revenue. Leveraging TikTok data, we analyze optimal traffic management strategies across different scenarios and delve into platform algorithm design to identify thresholds for optimal recommendations, aligning with platform objectives. Two central research questions guide our analysis: optimizing temporal traffic management from the streamer's perspective and enhancing platform algorithm design. Our first analysis centers on the optimal traffic management strategy for live streamers. Depending on the initial traffic level of the live stream, we identify three distinct scenarios. In the first scenario, we encounter streamers with nascent or marginal popularity, operating in smaller live rooms. These streamers engage in a strategy best described as "survive in the cracks" due to insufficient initial audience. Their focus is to maintain their current viewer base, given the slim prospects of winning the algorithmic promotion. The second scenario involves streamers positioned just below the threshold for receiving algorithmic promotion. These streamers employ more aggressive strategies to reach the threshold at an early stage of the livestream. In the third scenario, we observe streamers with substantial initial viewership who already benefit from algorithmic promotion and thus focus on sustaining their audience base. Interestingly, we find that the cash incentive is highest in the second scenario when the streamer is about to reach the algorithmic recommendation.

4 - Slide or Feed: Content Recommendation with Creators' Incentives

Shu Zhang, the Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Hongfan(Kevin) Chen, Renyu Zhang, Zhou Zhou

In online video platforms, optimizing recommendation systems is paramount for maximizing profit and incentivizing content creators.

Our work delves into the comparative efficacy of two modes of recommendation systems: the single-column (slide) and double-column (feed) formats. We aim to identify the circumstances under which each mode outperform the other in profit maximization and creator incentivization.

We build up a gaming model to describe the dynamics between recommendation modes and their impact on platform economics, where the two modes leads to differences in recommendation accuracy as well as user consumption cost. We illustrate how the performance of recommendation system is affected by both the modes and the total demand on the platform. The equilibrium results show that the strength of the recommendation algorithm plays an important role in this comparison.

Insights from our research contribute to a deeper understanding of how modes of recommendation system influence user content consumption patterns, creator engagement and the financial sustainability of online video platforms.

MB20

Summit - 340

Equilibrium in Complicated Games and Applications II

Invited Session

Decision Analysis Society

Co-Chair: Jeff Shamma, University of Illinois Urbana-Champaign, Champaign, IL, United States

1 - A review of terminologies and approaches regarding interdependencies in Project Portfolio Selection

Gustavo Vieira, Instituto Tecnologico de Aeronautica, São José dos Campos - SP, Brazil, Hévilla Oliveira, Jonatas Almeida, Carmen Belderrain

Project Portfolio Selection (PPS) is a relevant activity in organizations to define the best possible set of projects to pursue their strategic objectives. Although many authors recognize the importance of interdependencies between projects in PPS and program implementation, most approaches consider projects independently. In addition, there are different understandings of the meaning and the implications of interdependencies in the literature. Therefore, we present a literature review focused on project interdependencies within project portfolios. The main goals for this review were: (i) map the different definitions of the relationships between projects and their implications for PPS; (ii) summarize the several methods that address project interdependency in PPS; (iii) identify possible gaps for methodological improvement focused on optimizing project portfolios and enhancing program implementation by considering project interdependencies. Our search identified 72 papers from Scopus and Web of Science databases that were adherent to the parameters of a pre-defined protocol tailored to match the previously-mentioned goals. Hence, we depict an unprecedented contribution to PPS research by proposing a novel definition for relationships between projects that organizes and clarifies the meanings of the pre-existing definitions for interdependencies, interactions, and synergy and fulfills the existing gaps in those current definitions. We also present an overview about how future developments can improve

PPS by considering these interdependencies. Thus, this work summarizes the recent advances and future paths for dealing with PPS and improving PPS maturity within organizations by considering interdependencies between projects.

2 - AI based algorithm to compute Nash equilibrium in Bayesian games with continuous state and actions

Lichun Li, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Andres Eduardo Ho Lee

Bayesian Games are characterized by players possessing private information, such as the cost per unit for each player. This study specifically addresses scenarios where both private information and player actions are drawn from a continuous set. Previous research has examined the existence and uniqueness of a Bayesian Nash Equilibrium (BNE) in such games. We introduce an efficient computation method for this equilibrium using neural networks and game-oriented energy functions.

To compute the BNE, we begin by constructing a strategy model, represented as a neural network, for each player. Then, an energy function is developed to quantify the extent to which each player seeks deviation from the actions computed by their respective strategy models. By minimizing the energy function to zero, the strategy models converge towards the BNE of the game. This indicates that no player has an incentive to deviate from the actions determined by their neural network, given that other players rely on their respective strategy models for computing their actions.

The concept is validated by experimentation in symmetrical and asymmetrical multi-player rent-seeking games. Our proposed method computes a similar strategy that aligns closely with existing approaches found in the literature. Additionally, this method can seamlessly scale to accommodate scenarios involving more than two players because complexity increases with respect to the number of training samples instead of the number of players, requiring adjustments in the number of training samples as the player count increases.

3 - A Game-Theoretic Approach for Distributed Data Placement with Non-metric Access Costs

Rasoul Etesami, University of Illinois Urbana-Champaign, Champaign, IL, United States

Motivated by applications in web caches and content delivery in peer-to-peer networks, we consider the non-metric data placement problem and develop distributed algorithms for computing or approximating its optimal solutions. In this problem, the goal is to store copies of the data points among a set of cache-capacitated servers to minimize overall data storage and clients' access costs. We first show that the non-metric data placement problem is inapproximable up to a logarithmic factor. We then provide a game-theoretic decomposition of the objective function and show that a natural type of Glauber dynamics in which servers update their cache contents with probability proportional to the utility they receive from caching those data will converge to an optimal global solution for a sufficiently large noise parameter. In particular, we establish the polynomial mixing time of the Glauber dynamics for a certain range of noise parameters. Such a game-theoretic decomposition not only provides a good performance guarantee in terms of content delivery but also allows the system to operate in a fully distributed manner, hence reducing its computational load and improving its robustness to failures. Moreover, we provide another auction-based distributed algorithm, which allows us to approximate the optimal solution with a performance guarantee that depends on the ratio of the revenue vs. social welfare obtained from the underlying auction.

4 - Decentralized Strategies for Network Jamming Games

Brandon Van Over, Colorado State University, Fort Collins, CO, United States, Ali Pezeshki, Edwin Chong

We consider a two-player game over a multihop wireless network, involving multiple flows of packets. Each flow has its own source and destination, and each packet in a flow has a hard deadline. Player 1 wishes to determine a power control and routing policy to maximize the expected value of the average number of packets delivered prior to their deadlines, subject to a set of average power constraints. This performance metric is called timely throughput. Player 2 is a jammer that wishes to minimize the timely throughput by controlling the probability of failure for packet transmission across links in the network through injecting interference in the links, subject to its own set of average power constraints. We show that each player can employ a decentralized stationary randomized policy to play optimally against any fixed history-dependent randomized policy of their opponent. We investigate the convergence of such decentralized strategies to the Nash equilibrium of the game. Our work extends the results of Singh and Kumar (2018) on decentralized network control to network jamming games.

MB21

Summit - 341

Equilibrium in Complicated Games and Applications

Invited Session

Decision Analysis Society

Chair: Lichun Li, FAMU-FSU College of Engineering, Tallahassee, FL, United States

Co-Chair: Jeff Shamma, University of Illinois Urbana-Champaign, Champaign, IL, United States

1 - Multi-Agent Learning vs Equilibrium in Games

Jeff Shamma, University of Illinois Urbana-Champaign, Champaign, IL, United States, Sarah Toonsi

The framework of multi-agent learning in games explores the dynamics of how individual agent strategies evolve in response to the evolving strategies of other agents. Of particular interest is whether agent strategies converge to well-known solution concepts, such as Nash Equilibrium (NE) or its variants, under natural assumptions on the learning strategies. This talk begins with a brief introduction of the learning in games framework and an overview of both positive and negative results that describe when specific learning algorithms do or do not converge to NE, respectively. The talk concludes with a presentation of recent results that illustrate how a system theoretic approach to learning in games can leverage feedback control concepts (e.g., decentralized stabilization, strong stabilization, and robustness) to lead to new perspectives on what is or is not achievable in learning in games.

2 - Learning-Based Incentive Design for Cyberphysical Systems

Roy Dong, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Jung-Hoon Cho, Siqi Du, M. Umar B. Niazi, Cathy Wu

Incentive design for cyberphysical systems has the potential for significant societal impact. For example, in ground transportation, if we were able to incentivize eco-driving behaviors amongst drivers, we could reduce vehicle emissions between 10-45%. However, in these cyberphysical systems, the underlying physical dynamics can significantly complicate the incentive design problem: if one agent chooses a particular action, we not only have to consider how that action instantaneously affects other agents, but how that action (and resulting responses) propagate throughout the dynamics of the system. If we were able to explicitly account for this coupling in the form of a utility function, then classical incentive design optimization methods could be applied… but such a utility function is often not available in practice when the dynamics and coupling are complicated. In our work, we explore how we can abstract out the incentive design problem into a choice of 'driving policy' for each driver, rather than an explicit action. This lends itself to more amenable methods of simulation to determine the coupling between agents and their optimal responses when we can only access the dynamics through simulation or real-world deployment. We explore the efficacy of this method in the context of transportation networks, and demonstrate how these algorithms can automatically uncover the 'influential' agents who may most effectively be incentivized to steer the overall system. (For example, on a one-lane road, incentivizing the front car to slow down will implicitly force the cars behind it to slow down.)

3 - Learning Nash in Constrained Markov Games with a Near-Potential

Ceyhun Eksin, Texas A&M University, College Station, TX, United States, Soham Das

We develop a best-response algorithm for solving constrained Markov games assuming limited violations for the potential game property. The limited violations of the potential game property mean that changes in value function due to unilateral policy alterations can be measured by the potential function up to an error α. We show the existence of stationary ϵ-approximate constrained Nash policy whenever the set of feasible stationary policies is non-empty. Our setting has agents accessing an efficient probably approximately correct solver for a constrained Markov decision process which they use for generating best-response policies against the other agents' former policies. For an accuracy threshold $\epsilon>4\alpha$, the best-response dynamics generate provable convergence to ϵ-Nash policy in finite time with probability at least $1-\delta$ at the expense of polynomial bounds on sample complexity that scales with the reciprocal of ϵ and δ.

4 - Collaborative Coalitions in Multi-Agent Systems

Jason Marden, University of California, Santa Barbara, Santa Barbara, CA, United States, Bryce Ferguson

The emergence of new communication technologies allows us to expand our understanding of distributed control and consider collaborative decision-making paradigms. With collaborative algorithms, certain local decision-making entities (or agents) are enabled to communicate and collaborate their actions with one another to attain better system behavior. By limiting the amount of communication, these algorithms exist somewhere between centralized and fully distributed approaches. In this talk, we explore this collaborative paradigm and identify potential opportunities associated with collaborative decision-making in distributed resource allocation problems. In particular, we will provide a characterization for how the level and structure of collaboration impacts the efficiency of the emergent collective behavior. Note that having such a characterization is essential for identifying whether or not the communication cost necessary to facilitate the desired collaborations is worth the gains in system welfare.

5 - Enhancing Social Welfare Leveraging Pure Nash Equilibrium in Integer Programming Games for Aquatic Invasive Species Prevention

Hyunwoo Lee, Virginia Tech, Blacksburg, VA, United States, Robert Hildebrand, Esra Toy

This study introduces a framework designed to enhance social welfare within non-cooperative integer programming games. With advancements in the IPG literature, attaining the best Pure Nash Equilibrium in terms of social welfare has become feasible. Recognizing the best PNE as both a fair and optimal equilibrium, we propose two integer programming models that leverage this equilibrium for post-game refinement: a refined social welfare model and a refined coalitional model. Both models ensure each player's utility at the socially optimum PNE is preserved while aiming to enhance the overall social welfare. The first model maintains the original non-cooperative game settings, preserving individual player strategies. The second model pools strategies within coalitions, promoting collaborative benefits. We apply this framework to edge-weighted budgeted maximum coverage games for Aquatic Invasive Species prevention, aligning player objectives with both state-level and county-level goals. Computational experiments demonstrate significant improvements in social welfare achieved through our proposed methods.

MB22

Summit - 342

Data-Driven Excellence in Healthcare: Merging Optimization, Stochastic Decision-Making, and Operations Management

Panel Session

Decision Analysis Society

Co-Chair: Mohammad Alipour Vaezi, Virginia Tech, BLACKSBURG, VA, United States

1 - Moderator Panelist

Mohammad Alipour Vaezi, Virginia Tech, BLACKSBURG, VA, United States

2 - Panelist

Soroush Saghafian, Harvard University, Cambridge, MA, United States

3 - Panelist

John Birge, University of Chicago, Chicago, IL, United States

MB23

Summit - 343

Building MCDM/A Models: Practical and Methodological Issues

Invited Session

Multi Criteria Decision Making

Chair: Eduarda Frej, Universidade Federal de Pernambuco, Recife, Brazil

Co-Chair: Danielle C. Morais, Universidade Federal de Pernambuco - UFPE, Recife - PE, Brazil

1 - Supporting Resource Allocation to Combat Drought Through a Multicriteria Sorting Model

Danielle C. Morais, Universidade Federal de Pernambuco - UFPE, Recife - PE, Brazil, Gabriel Castro

This study addresses the global impact of drought on water resources and aims to support resource allocation to combat its adverse effects. It proposes a multicriteria sorting model to assist the government in evaluating municipalities in a river basin in Brazil. The model categorizes the municipalities into high, moderate, and low-priority tiers for effective drought mitigation efforts. The findings demonstrate the model's capacity to support strategic public policies, enabling informed resource allocation and the development of targeted actions to address the specific needs of cities most impacted by drought, thereby mitigating the overall adverse effects of this natural disaster.

2 - Analytical Exploration of the Theoretical Foundation of Fitradeoff Method

Jonatas Almeida, Universidade Federal de Pernambuco , Caruaru, Brazil, Eduarda Frej, Lucia Roselli, Adiel Almeida

The ability to make effective decisions is fundamental to personal, professional, and organizational success, as it determines the course of actions to be followed, influencing the results obtained through the use of efforts and resources. Decision-making becomes particularly challenging when the decision maker has multiple objectives, requiring judgments to express how much they are willing to sacrifice on one objective to enhance another. The FITradeoff method, belonging to the set of Multi-Attribute Value Theory (MAVT) methods, is based on the tradeoff procedure and aims to analyze how the subspace of weights, limited by partial information, defines values and preference relationships between alternatives to construct a recommendation. Methodological improvements to FITradeoff have been published in various journals and conferences, including expansions to address different problematics, improved usability and cognitive performance through neuroscience studies, and the incorporation of holistic evaluation in the elicitation process combined with the decomposition procedure. This study seeks to analyze how the combined elicitation procedure is used to explore the weight subspace, how it can help identify inconsistencies in the decision maker's statements, and how certain heuristics can be employed to expedite the reduction of the viable weight subspace, aiding in the more efficient construction of a recommendation.

3 - A Multicriteria Decision Model for Carrying Company Selection in the Clothing Sector in Brazil

Layra Almeida, Universidade Federal de Pernambuco, Caruaru, Brazil, Jonatas Almeida, Eduarda Frej

The clothing industry is an important sector of Brazilian economy. Clothing companies are distributed in several regions of the country. These companies interact through products and services. Brazil is a country with continental dimensions, and carrying is an essential service for integrating the supplier chain. In this context, a multicriteria decision model was developed for a carrying company selection in the Brazilian clothing sector. A framework for building multicriteria decision models was applied. The framework has a successive refinement process, which allows for enriching the decision model by returning to any previous step to incorporate more information. Definition of basic elements of the problem and methodology are conducted through the application of the framework. A well-structured multicriteria decision model presents advantages as a proper representation of a real problem, and provides a robust and reliable recommendation for the addressed decision problem.

4 - Choice of Method for Multicriteria Decision Using a Framework for Structuring and Building Models

Adiel De Almeida, Universidade Federal de Pernambuco, Recife, Brazil, Lucia Roselli, Eduarda Frej, Danielle C. Morais

The Choice of method for multicriteria is discussed in this presentation. This choice can be done in several ways. This is one of the main issues in the decision process, which uses a Framework for structuring the decision problem and then building a multicriteria decision model. The preference modeling is the main focus on the choice of MCDM/A methods, in which the main steps is the analyzes of compensatory and non-compensatory preferences. An information and decision support system has been developed to assist the Analysist and the Decision Maker in this process.

5 - A Structured Protocol for Sensitivity Analysis in the Fitradeoff Method

Eduarda Frej, Universidade Federal de Pernambuco, Recife, Brazil, Pedro Henrique De Souza, Wagner Gouveia, Adiel de Almeida

This work presents a structured protocol to perform Sensitivity Analysis in the FITradeoff method, in order to verify the robustness of the solutions obtained. Sensitivity Analysis is a crucial step in Multicriteria Decision Making/Aiding models, in order to provide to the DM a view on how robust or sensitive are the solutions obtained by the applied method, considering variations on inputs' values. However, performing a sensitivity analysis can be challenging in partial information methods, as is the case of the FITradeoff method, since partial rankings of the alternatives that may be obtained. The proposed protocol has basically two steps. First, af a robustness analysis of the original solution is perfomed based on a Monte Carlo simulation process, and then the application of the Kendall Correlation Statistical Test is made, to verify the statistical significance of the correlations between the rankings obtained in the simulation analysis compared to the original ranking. The proposed protocol was embedded into the FITradeoff Decision Support System, enabling DMs to have a better understanding of possible variations on the solution obtained and further practical implications.

MB24

Summit - 344

Advancements in Robotics and AI Applications

Contributed Session

Chair: Ranit Senapati, IIT KHARAGPUR, 721302

1 - Exploring the Interplay between Brand Identity and AI-Based Service Robots

Dongtan Li, Auburn University, Auburn, AL, United States, Pei Xu

With the rapid advancement of artificial intelligence technology, generative AI has dramatically attracted abundant attentions in both academia and industry. AI-based service robot, as one of the most creative and commercial applications, has made great contributions to increase service efficiency and reduce labor cost, especially in the hospitality industry. Recent research indicated a positive correlation between AI service robots and customer satisfaction ratings in hospitality (Borghi et al., 2023); whereas studies on brand identity in terms of brand personality are still limited. As a consequence of COVID 19 pandemic, consumers are increasingly aware of the risk of infectious diseases from interpersonal interactions and the implementation of service robot has been significantly boosted in the hospitality industry. However, certain utilizations of service robot can also precipitate the transformation of brand identity, which may either strengthen or damage the brand image. In order to investigate the interplay between brand identity and service robots, this study suggests brand personality as the main measurement of brand identity, which describes a brand from five dimensions including sincerity, excitement, competence, sophistication, and ruggedness (Aaker 1997). This study further examines how the implementation of service robots impact brand personality in the short- and long-term, and how such impact is moderated by brand position. Findings of our study can help hospitality brands better recognize the possible alteration of their brand identity while deploying AI service robot.

2 - Analyzing restaurant customers' and owners' preferences for serving robots using choice experiment and latent class analysis

Changeun Park, Kyung Hee University, Yongin, Korea, Republic of, Jungwoo Shin

Serving robots are growing by innovating customer experiences, providing positive benefits to both restaurant customers and owners by improving productivity, resolving labor shortages, and providing new experiences and services. However, along with digital literacy, there are also negative perceptions about robots, such as the discomfort of robots resembling human and human replacement. The purpose of this study was to identify serving robot types and restaurants operation types that restaurants customers and owners prefer using Choice Experiments (CE) and Latent Class Analysis (LCA). The study found that customers were divided into two groups: technology change adopters, who would accept supporting both voice and app calls with serving robots, and technology change avoiders, who would not accept serving robots and would prefer only voice calls, while owners were divided into two groups: those who valued interactivity, such as the call method and those who valued the anthropomorphism of the robot more than the call method. The analysis of the acceptance scenarios by attribute for each group showed that both groups preferred a combination of serving robots and human labor, with customers preferring non-humanoid serving robots and owners preferring fixed type or humanoid serving robots. Customers were least likely to accept humanoid serving robots, suggesting the possibility of an uncanny valley effect, and restaurants owners should consider it when introducing serving robots.

3 - Minimizing energy consumption by controlling the base positions of a robotic arm based on a statistical approach

Seung Hyeon Eom, Kyung-Hee University, Yongin-si, Korea, Republic of, Hyun Woo Jeon

The usage of industrial robots, especially robotic arms, is in an increasing trend. Reducing the energy consumption of industrial robotic arms is one of the most important energy-related issues in the industrial sector. Although the base position of robotic arms is crucial for managing and saving energy consumption, this research problem has not been studied comprehensively. Moreover, rather than more scientific research or data, the intuition or experience of engineers is used as the basis to determine the base position. Therefore, this study aims to optimize the base position of robotic arms by minimizing energy consumption for tasks. For the optimization formulation, we use robotic arm power models estimated from a statistical method, in which we consider three simple input variables: starting point, ending point, and payload. We also show the convexity of the objective function (energy consumption for a given task). From analysis, we present that the best base position of robotic arms for each task can be determined by minimizing energy consumption.

4 - A critical review on applications of Surrogate Modeling, Machine Learning and Deep learning techniques in New Product Development

Ranit Senapati, Indian Institute of Technology Kharagpur, Kharagpur, India, Biswajit Mahanty

The application of advanced computational techniques such as surrogate modeling, machine learning, and artificial intelligence has enhanced decision making and innovation in the domain of new product development. Designers can easily construct surrogate models from simulations or experiments, enabling them to evaluate design alternatives, optimize product performance, and explore the design space efficiently. Additionally, these surrogate models serve as invaluable tools for sensitivity analysis and uncertainty quantification, aiding in risk management and robust design. Machine learning techniques enable designers to perform predictive modelling, pattern recognition, and anomaly detection, facilitating data-driven decision making at every product development lifecycle stage. Through techniques such as natural language processing, computer vision, and deep learning, AI systems can analyze unstructured data, extract actionable insights, and generate innovative ideas, revolutionizing ideation and concept generation processes. In conclusion, the synergistic integration of surrogate modeling, machine learning, and artificial intelligence heralds a new era of agility, innovation, and competitiveness in product development. By harnessing the predictive power of these technologies, organizations can expedite time-to-market, mitigate risks, and deliver customer-centric products that resonate with evolving market demands. However, addressing challenges related to data quality, algorithmic biases, and organizational readiness is imperative to realize the full potential of these computational techniques.

MB25

Summit - 345

Hexaly/Artelys

Invited Session

Technology Showcase

1 - Hexaly, a New Kind of Global Optimization Solver

Fred Gardi, Hexaly, Brooklyn, NY, United States

Hexaly is a new kind of global optimization solver. Its modeling interface is nonlinear and set-oriented. It also supports user-coded functions, thus enabling black-box optimization and, more particularly, simulation optimization. In a sense, Hexaly APIs unify modeling concepts from mixed-linear programming, nonlinear programming, and constraint programming. Under the hood, Hexaly combines various exact and heuristic optimization methods: spatial branch-and-bound, simplex methods, interior-point methods, augmented Lagrangian methods, automatic Dantzig-Wolfe reformulation, column and row generation, propagation methods, local search, direct search, population-based methods, and surrogate modeling techniques for black-box optimization.

Regarding performance benchmarks, Hexaly distinguishes itself against the leading solvers in the market, like Gurobi, IBM Cplex, and Google OR Tools, by delivering fast and scalable solutions to problems in the space of Supply Chain and Workforce Management like Routing, Scheduling, Packing, Clustering, Matching, Assignment, and Location problems. For example, on notoriously hard problems like the Pickup and Delivery Problem with Time Windows or Flexible Job Shop Scheduling with Setup Times, Hexaly delivers solutions with a gap to the best solutions known in the literature smaller than 1% in a few minutes of running times on a basic computer.

In addition to the Optimizer, we provide an innovative development platform called Hexaly Studio to model and solve rich Vehicle Routing and Job Shop Scheduling problems in a no-code fashion. The user can define its problem and data, run the Optimizer, visualize the solutions and key metrics through dashboards, and deploy the resulting app in the cloud – without coding.

2 - Model and solve nonlinear optimization problems with Artelys Knitro

Richard Waltz, Artelys, Los Angeles, CA, United States

The challenges faced by operational research (OR) practitioners, whether in operational scenarios or during techno-economic analyses, are typically intricate and inherently nonlinear. A skilled modeler's task is to simplify this complexity and identify the core issue whose resolution untangles the rest.

A common strategy is to linearize the problem to leverage the wealth of theoretical and practical tools provided by linear programming, such as duality, sensitivity analysis, and efficient algorithms. While this approach can be effective, one can often obtain significantly better solutions by directly solving a nonlinear model that more closely resembles the actual problem.

Artelys Knitro is the leading solver focused on large-scale, nonlinear (potentially non-convex), optimization problems. Knitro offers both interior-point and active-set algorithms for continuous models, as well as tools for handling problems with integer variables and other discrete structures. This tutorial will introduce the key features of Knitro, and demonstrate how to use Knitro to model and solve optimization problems in various environments.

This tutorial will showcase several instances of nonlinear problems, typically addressed through linear relaxation, yet warranting a direct approach. It will delve into the methodologies and tools employed to tackle these challenges. This tutorial aims at an audience familiar with the basics of mathematical optimization, focusing on practical examples.

MB26

Summit - 346

Academic Job Search Panel

Panel Session

Job Placement Services

Co-Chair: Priyank Arora, University of South Carolina, Columbia, SC, 29208, United States

1 - Moderator Panelist

Priyank Arora, University of South Carolina, Columbia, SC, United States

2 - Panelist

Susan Hunter, Purdue University, West Lafayette, IN, United States

1

3 - Panelist

Tugce Isik, Clemson University, Clemson, SC, United States

2

4 - Panelist

Qiuping Yu, Georgetown University, Washington, DC, United States

4

5 - Panelist

Sina Khorasani, Vanderbilt University, Franklin, TN, United States

MB27

Summit - 347

TIMES Best Dissertation Award 2024

Award Session

Technology, Innovation Management and Entrepreneurship

Chair: Janne Kettunen, The George Washington University, Vienna, VA, United States

1 - Innovation Management on Online Platforms

Sidika Tunc Candogan, University College London, London, United Kingdom

In my thesis, I study operational decisions on online platforms that facilitate crowd-based innovation and new product development, and I provide insights into how the economic value generated on these platforms can be enhanced. In the first two chapters, I focus on crowdsourcing platforms (e.g., InnoCentive and Topcoder) that create value by regularly organizing innovation contests to find solutions to their clients' innovation-related problems from the crowd. In the third chapter, I focus on crowdfunding platforms (e.g., Kickstarter and Indiegogo), where entrepreneurs launch campaigns to crowdsource funding to bring their innovative ideas to life. In this talk, I will briefly talk about what motivated me to work on the project in each chapter, research questions, key tradeoffs, and managerial insights.

2 - Emerging Technologies in Healthcare Operations

Jiatao Ding, Nanyang Technological University, Singapore, Singapore

This dissertation investigates the operational implications of emerging technologies in healthcare through three essays. The first essay examines strategic collaborations between biopharmaceutical firms and technology providers in acquiring analytics capabilities. Using event study methodologies, we find that such collaborations yield significant positive abnormal returns, especially when aimed at enhancing prelaunch R&D or involving multiple therapeutic areas. We also explore how firm characteristics influence the impact of these collaborations on shareholder value. The second essay focuses on malaria-endemic countries, analyzing optimal intervention strategies with subsidy program and traceability technology adoption for combating counterfeit antimalarial drugs. Using a game-theoretic model, we demonstrate that traditional purchase subsidies may not be effective in the presence of uncertified distribution channels. Instead, we advocate for tailored approaches, supported by numerical analysis, to design effective policies and technologies for improving access to life-saving medicines. The third essay assesses the operational implications of virtual triage chatbot adoption in acute care settings, addressing the challenges of patient self-triage inaccuracy. Using a queueing game model, we explore the trade-offs between informativeness and volume in virtual triage recommendations, highlighting the need for custom solutions and effective regulation to enhance acute care delivery. Overall, these essays contribute actionable insights for policymakers, practitioners, and researchers, emphasizing the importance of strategic collaborations, tailored interventions, and customized technology adoption in advancing healthcare operations and enhancing patient outcomes.

3 - Managing the nanostore supply chain: Base-of-the-pyramid retail in emerging markets

Rafael Escamilla, Arizona State University, Tempe, AZ, United States

The presentation will delve into the intricacies of nanostore supply chains in emerging markets, focusing on the financial, material, and information flows within them. First, I will discuss how suppliers benefit operationally from allowing short-term payment delays, despite inherent risks. The discussion will then shift to the challenges of digitization for nanostore shopkeepers, who fear that adopting digital technologies could expose tax evasion. Then, I will discuss the competition between nanostores and convenience stores, highlighting the benefits of value-added digital services. Finally, I will discuss the strategic decision suppliers face in determining their visit frequency, which directly impacts shopkeeper behavior.

4 - Information Sharing and Operational Transparency on On-Demand Service Platforms

Swanand Kulkarni, Tuck School of Business, Dartmouth College, Hanover, NH, United States

The three chapters in this dissertation examine the operational practices of on-demand service platforms, pertaining to information sharing and operational transparency on the supply-side. On-demand service platforms such as ridesharing, food delivery, grocery delivery, and courier delivery platforms critically depend on the services of workers, who are independent contractors. Given that these workers have discretion over their labor supply decisions, the platform's information provision to workers plays a key role in influencing workers' decisions and in eventually meeting the customer demand. In this dissertation, I employ game-theoretic modeling and conduct incentivized experiments with human subjects to evaluate the implications of a platform's practices around information sharing and operational transparency with workers, for workers' decisions and for the platform's operational outcomes. In Chapter 1, using game-theoretic modeling and lab experiments, we investigate how an on-demand service platform's mechanism to share demand–supply mismatch information spatially affects drivers' relocation decisions and the platform's matching efficiency. Chapters 2 and 3 experimentally examine how a platform's compensation practices, considered jointly with the level of transparency that workers have into the pay determination process, influence worker participation with the platform. In Chapter 2, we examine the impact of the intuitiveness and transparency of a platform's compensation model on worker participation. In Chapter 3, motivated by platforms' experiments around managing the commission that they charge workers, we study the impact of consistency in (across service instances) and transparency of the platform's commission on worker participation.

5 - Three Essays on New Product Development: Creating Value from Internal and External Innovation

Hossein Nikpayam, Frankfurt School of Finance & Management, Frankfurt, Germany

Innovation drives business success, particularly through new product development (NPD). This dissertation explores the intricate balance between resource allocation, communication strategies, and incentives in managing NPD initiatives. It investigates how firms can optimize the allocation of scarce resources, highlighting the importance of balancing flexibility with dedicated resources to enhance project selection without excessive information acquisition costs. Additionally, it examines the role of strategic communication and financial incentives in guiding resource allocation, emphasizing how these factors influence decision-making in environments with varying levels of internal and external project information. Finally, the research delves into the dynamics of innovation contests, identifying conditions under which truthful and fully informative performance feedback is most effective. These insights collectively offer practical strategies for firms to effectively navigate the complexities of innovation management.

MB28

Summit - 348

Waste and the Circular Economy

Invited Session

MSOM: Sustainable Operations

Chair: Adem Orsdemir, University of California Riverside, Riverside, CA, United States

Co-Chair: Anna Saez de Tejada Cuenca, IESE Business School, Barcelona, Spain

1 - Competitive mass customization and environmental sustainability in fashion

Adem Orsdemir, University of California Riverside, Riverside, CA, United States, Aydin Alptekinoglu

We explore adoption of mass customization (MC) -- in place of mass production (MP) -- in a profit-driven competitive environment as a solution to the serious problem of overproduction (and all the environmental waste that it generates) endemic in the fashion industry. Building on a demand model sensitive to product variety, price and a lead time for mass-customized products, we analyze a duopoly competition game that includes production technology selection between MP and MC. We then take the equilibrium of this game as a characterization of the profit-driven market outcome, and analyze its environmental impact. This analysis teaches us when adoption of MC leads to lower environmental impact and why. We uncover, for example, when MC adoption by one firm (with the other sticking with MP) leads to higher profits for both firms and lower total environmental impact -- a so called win-win-win. Also, generally speaking, low-to-medium marginal cost of production is more conducive for MC adoption to lead to greener outcomes. This is good news in that lower marginal costs typically imply lower prices and higher sales volumes in practice, pointing to the part of the fashion industry that has larger absolute environmental impact.

2 - Lowering the environmental impact of fashion retail assortments: Balancing profit maximization and impact minimization

Jean-Sebastien Matte, McGill University, Montreal, QC, Canada, Mehmet Gumus, Robert Rooderkerk

There is a general consensus that the fashion industry needs to reduce its environmental impact. However, there is confusion and misunderstanding between retailers, consumers, and policymakers as to how the different dimensions of environmental impact in fashion are understood and valued. In addition, there are currently no regulation in place to curb the industry's highly impactful production.

We propose an integrated framework that combines an incentive-aligned choice-based conjoint (IA-CBC) experiment, a Hierarchical Bayes Multinomial Logit (HB-MNL) model, and joint product-line and pricing optimization model to measure preferences and willingness-to-pay (WTP) for three environmental attributes---Recycled Content, Durability, and Circularity Design---and the trade-off between profits and environmental constraints (regulations). Our experimental design also includes a between-subjects manipulation to study the impact of priming consumers as a reconciling mechanism between consumers, retailers and policymakers.

We find that environmental attributes are important, but that Price remains the main driver of choice. Our results also define a clear hierarchy of importance between environmental attributes and their levels, and highlight confusion towards Durability. We find that constraints on Recycled Content have the most impact on business performances, and that it is almost always optimal for the retailer to offer a Circularity Design level.

Finally, we find that priming consumer significantly shifts attribute importance, increases WTP, and consequently considerably eases the profit-environmental impact trade-off.

For managers, our methodology and results provide a strategic decision support tool to optimally adapt assortments to consumer demand and regulations. For policymakers, we provide valuable empirical evidence to develop fair regulations.

3 - Customer-to-Customer Returns Logistics: Can it Mitigate the Negative Impact of Online Returns?

(Ayse) Sena Eruguz, Vrije Universiteit Amsterdam, AMSTERDAM, Netherlands, Oktay Karabag, Eline Tetteroo, Carl van Heijst, Wilco van den Heuvel, Rommert Dekker

Customer returns are a major problem for online retailers due to their economic and environmental impact. In this talk, we introduce a new concept for handling online returns: customer-to-customer (C2C) returns logistics. The idea behind the C2C concept is to deliver returned items directly to the next customer, bypassing the retailer's warehouse. To incentivize customers to purchase C2C return items, retailers can promote return items on their webshop with a discount. We build the mathematical models behind the C2C concept to determine how much discount to offer to ensure enough customers are induced to purchase C2C return items and to maximize the retailer's expected total profit. Our first model is a customer-based formulation of the problem and provides an easy-to-implement constant-discount-level policy. Our second model formulates the real-world problem as a Markov decision process (MDP). Since our MDP suffers from the curse of dimensionality, we resort to simulation optimization and reinforcement learning methods to obtain reasonably good solutions. We apply our methods to data collected from a Dutch fashion retailer. Our results indicate that significant benefits can be achieved with the C2C returns program in terms of both expected profit and return rate. Even in cases where the cost-effectiveness of the C2C returns program is not pronounced, the proportion of customer-to-warehouse returns to total demand becomes lower. Therefore, the system can be defined as more environmentally friendly. The C2C concept can help retailers financially address the problem of online returns and meet the growing need for reducing their environmental impact.

4 - Timing of remanufactured product introductions over evolving product generations in steady state

Ana Muriel, University of Massachusetts, Amherst, MA, United States, Yue Jin

We analyze alternative remanufacturing strategies for a firm that introduces new product generations at regular intervals. The literature has mainly focused on steady state models where the new product competes with a certain remanufactured version of the product each interval. We propose a new steady state model that breaks the generation interval into two periods. The new product is introduced at the beginning of period 1 and competes with the remanufactured version of the old generation. At some point, remanufacturing the new generation product will become feasible and the company can choose to transition to period 2, offering remanufactured products of the new generation which is more appealing to the customer. We quantify the benefits associated with the two-period strategy, analyze the timing decision for offering the new version of the remanufactured product, characterize structural properties of the solutions, and study their sensitivity to various model parameters.

5 - Achieving Resilience Amidst Crises in U.S. Dairies: The Importance of Sustainability Investments and Women in Farming

Veronica Villena, Arizona State University, Tempe, AZ, United States, Elizabeth Ransom, David Abler

The COVID-19 pandemic created major supply chain disruptions, forcing managers to focus their attention on short-term survival. Using the dairy sector, we show that investing in long-term sustainability practices builds resilience. This study suggests that now is the time to reinforce sustainability investments in order for dairy farms to better navigate future disruptions. Building upon the research stream on sustainability and resilience in agriculture operations, we conducted two empirical consecutive studies. The first, a qualitative study, reveals (a) the types of sustainability interventions made by dairy farms, (b) the unknown challenges that dairy farms faced during COVID, and (c) the reasons why farms long investing in sustainability practices exhibited higher resilience. It also shows some unique differences between women and men operators. The second, a quantitative study, tested the propositions emerging from the first study in a large-scale survey involving 588 Pennsylvania dairy farms. Both studies show that farms that have long invested in environmental, labor, and animal welfare practices are more likely to continue their operations and implement an emergency plan that allows them to run safe operations. Interestingly, they reveal that, compared to environmental and animal welfare practices, dairy farms' human resource practices are very nascent and precarious, which is worrisome given this sector's severe labor shortage and high turnover rate. Furthermore, the results show that women operators benefit more from sustainability investments than men operators, even when the odds are against for them.

MB29

Summit - 420

Advances in network optimization under uncertainties: Theories and applications

Invited Session

OPT: Integer and Discrete Optimization

Chair: Thanakrit Piyachayawat, Texas Tech university, Lubbock, TX, United States

1 - Network Enhancement via Robust Interdiction Methods

Thanakrit Piyachayawat, Texas Tech university, Lubbock, TX, United States

In the classic network interdiction games, the leader (the first decision-maker) traditionally acts as an attacker, aiming to disrupt their opponent's solutions such as paths, connected components, or spanning trees. In this work, we propose a novel reinterpretation of the leader's role as a network user, focusing on enhancing network structures while mitigating uncertainties originating from multiple sources. Using new reformulation methods, we address this class of problems effectively. We demonstrate that, across a range of feasible decisions dictated by optimistic and pessimistic perspectives, optimal robust solutions can be obtained through single-level mixed linear programming formulations, which are tractable via proposed cut-generation approaches. Moreover, using facet analysis techniques, we show that our reformulated models simplify this network enhancement problem while preserving its optimal characteristics. These findings open up new avenues for studying and addressing more complex network problems.

2 - A Robust Bi-Level Network Interdiction with Applications in Human Trafficking Disruption

Daniel Lopes da Silva, Clemson University, Clemson, SC, United States, Thomas Sharkey, Yongjia Song

We investigate a robust bi-level network interdiction problem motivated by applications in human trafficking. In this problem, the follower, who operates the network, solves a minimum cost flow problem. The leader, who interdicts the network, minimizes the number of arcs from a special set with flow on them in the minimum cost flow solution obtained by the follower. The problem includes data uncertainty as the leader does not know the cost vector but only that it belongs to a polyhedral set. The follower has complete knowledge about its own parameters and therefore makes decisions in a ``wait-and-see'' fashion.

3 - Optimizing Microgrid Investment to Improve Resilience Using Stochastic Optimization

Ramsey Rossmann, University of Wisconsin — Madison, Madison, WI, United States, James R. Luedtke, Line Roald

Distribution networks in the power grid are rapidly changing with increases in distributed energy resources, posing challenges to serving load as well as opportunities for increasing reliability. Installing switches and power sources within a distribution feeder can increase operational flexibility of the feeder and allow for more load to be served during outages. In extreme circumstances, widespread outages can shut down feeders for days even when resources within the feeder could supply power to serve some load. A key challenge in many outages is that when a line has a fault (e.g., a tree falls on it), the fault will spread until it can be isolated, causing more parts of the feeder to be de-energized. To decrease the impact of outages, utilities may consider installing switches to isolate faults and change the feeder's topology as well as generators, batteries, and solar arrays to provide power for the feeder when the substation is unavailable. Given uncertainties in outages, load, and solar availability as well as the non-convexities of binary decisions and power flow constraints, optimizing such feeder investments is challenging. We propose a simplified model for making optimal investment decisions in a feeder under uncertainty by approximating feeder operations with a network flow formulation. We use various approximation and decomposition approaches to find solutions to the resulting two-stage stochastic integer program. Preliminary results suggest these models are good approximations of the original model and perform well across a variety of instances.

4 - Relay-Hub Network Design for Consolidation Planning Under Demand Variability

Onkar Kulkarni, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan, Benoit Montreuil

In this work, we study a two-stage stochastic program to design relay-based transportation network. In the first stage, we position the relay-hubs and size them, and design a minimum-cost consolidation plan to transport commodities upon realization of commodity demand uncertainty in the second stage. Due to commodity consolidation considerations, second stage consists of general integer decision variables which makes the problem challenging to solve. We devise a decomposition-based branch-and-cut algorithm with nested Benders decomposition and integer L-shaped method to solve the problem-at-hand exactly. To test and validate our methodology, we design large-scale resilient relay networks to be used for finished vehicle deliveries for a US-based car manufacturer that partnered with our research team. Finally, we provide managerial insights regarding inducing flexibility in the relay network to hedge against commodity demand variability.

MB30

Summit - 421

Quantum Computing and Optimization (Part 2)

Invited Session

OPT: Linear and Conic Optimization

Chair: Mohammadhossein Mohammadisiahroudi, Lehigh University, Bethlehem, PA, 18015, United States

Co-Chair: Zeguan Wu, Lehigh University, Bethlehem

1 - Solving Stochastic Binary Optimization Problems with Quantum Approximate Optimization Algorithm

Anthony Wilkie, University of Tennessee-Knoxville, Knoxville, TN, United States

There are many problems in stochastic programming that are currently intractable using today's classical computers. These problems usually involve a large number of binary variables with a recourse function that may be non-convex and discontinuous. Quantum computing is a new technology that has the potential to solve these problems. Specifically, the Quantum Approximate Optimization Algorithm (QAOA) is a quantum algorithm that has been shown to approximate solutions to combinatorial optimization problems. As a result, this project will explore the use of QAOA to solve stochastic binary programs. We propose a method to solve the deterministic equivalent of the stochastic program using QAOA. We will show how to encode our problem in a quantum circuit and how QAOA is able to approximate the solution. This proof of concept will be tested on small problems to demonstrate the potential of QAOA to solve stochastic binary programs. Then we explore potential methods to create more efficient quantum circuits and improve the performance of QAOA.

2 - Quantum Relaxation for Combinatorial Optimization Problems

Zichang He, JPMorgan Chase, New York, NY, United States, Rudy Raymond, Ruslan Shaydulin, Marco Pistoia

Quantum optimization holds promise for providing advantages in solving combinatorial optimization problems. One key component involves the exact encoding of a classical cost function into a quantum Hamiltonian, which is widely utilized in quantum algorithms. However, an alternative approach involves relaxed encoding, inspired by quantum random access codes. This quantum relaxation enables the handling of problems that are 2-3 times larger while utilizing the same number of qubits. Despite this, quantum algorithms for solving the relaxed Hamiltonian have received less attention. In this talk, we will discuss recent advancements in quantum algorithms and their implementation for addressing relaxed optimization problems.

3 - Constrained local Hamiltonians: quantum generalizations of Vertex Cover

Sankara Sai Chaithanya Rayudu, University of New Mexico, Albuquerque, NM, United States, Ojas Parekh, Kevin Thompson

Given the diversity of combinatorial optimization problems and their impact on the development of new algorithms, it is natural to ask if we can find quantum generalizations of some of the constrained combinatorial problems, which, in turn, can inspire new quantum algorithms. In this work, we propose a quantum generalization of the vertex cover problem, which we call 'Transverse Vertex Cover' (TVC). We prove that TVC is in the complexity class StoqMA-hard and develop an approximation algorithm for it based on a quantum generalization of the classical local ratio method. Additionally, we demonstrate our quantum local ratio method on a traditional unconstrained quantum local Hamiltonian version of the Vertex Cover, which is equivalent to the well-studied anti-ferromagnetic transverse field Ising model in physics.

4 - Iterative Refinement for Quantum Computing Algorithms

Mohammadhossein Mohammadisiahroudi, Lehigh University, Bethlehem, PA, United States, Tamás Terlaky, Brandon Augustino, Giacomo Nannicini, Zeguan Wu, Muqing Zheng, Pouya Sampourmahani

Quantum linear system algorithms (QLSAs) offer potential complexity advantages over classical methods for solving linear systems, with respect to dimension. Yet, extracting classical solutions requires a quantum tomography algorithm (QTA), which increases both error and time complexity. Recent advancements have integrated QLSA with QTA to develop quantum interior point methods (QIPMs) for solving different types of linear conic optimization problems. Although these innovations promise polynomial speed-ups relative to dimension, their efficiency is limited by polynomial dependencies on precision and the condition number, hindering their utility for finding a precise solution. Our work utilizes iterative refinement methods (IRMs) to achieve precise solutions by iteratively employing quantum oracles of limited precision. We demonstrate how IRMs can exponentially accelerate QLSA combined with QTA and QIPMs concerning precision. Furthermore, IRMs mitigate the impact of the condition number on the computational complexity of QIPMs. Another challenge for QLSAs is that their gate complexity is far larger than what is implementable on current noisy intermediate-scale quantum computers. We also explore the potential of IRMs in facilitating the early implementation of QLSAs, enhancing the practicality and precision of quantum computing methods in solving systems of linear equations and consequently, solving linear optimization problems.

5 - Lagrangian Reformulation for Nonconvex Optimization: Tailoring Problems to Specialized Solvers

Rodolfo Alexander Quintero Ospina, Lehigh University, Bethlehem, PA, United States, Juan Vera, Luis Zuluaga

Advancements in computing technologies, such as quantum and Ising devices, have led to a surge of interest in studying different ways to reformulate nonconvex optimization problems, especially those involving binary variables. In this talk, I will present results that show the equivalence between a class of equality-constrained nonconvex optimization problems and their Lagrangian relaxation, bridging classical optimization theory with emerging quantum computational approaches and filling a gap in classical optimization literature. Moreover, we explore the implications of these findings for solving Integer Programs and Mixed Integer Problems using quantum optimization algorithms.

MB31

Summit - 422

Interplay Between Optimization, Statistics, and Learning - Part II

Invited Session

OPT: Machine Learning

Chair: Lijun Ding, University of California San Diego, La Jolla, CA, 77843, United States

Co-Chair: Liwei Jiang, Georgia Institute of Technology, Atlanta, GA, United States

1 - *Non-Asymptotic Global Convergence Analysis of Bfgs*

Aryan Mokhtari, University of Texas at Austin, Austin, TX, United States

In this talk, we discuss the non-asymptotic global convergence rates of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for two scenarios: when it is implemented with an exact line search and when it is implemented with an inexact line search that utilizes the Armijo-Wolfe conditions. For both cases, we establish global convergence rates that hold for any initial point and any initial Hessian approximation. To the best of our knowledge, these are the first results that are able to characterize the overall iteration complexity of the BFGS quasi-Newton method.

2 - Clip-Ocular: Seeing Through the Lense of Visual Language Models

Mahdi Soltanolkotabi, University of Southern California, Los Angeles, CA, United States, Zalan Fabian, Berk Tinaz

Multimodal generative AI models have achieved unparalleled success as conversational agents capable of interpreting both image and text inputs. However, recent research has cast serious doubt on the visual understanding and reasoning capabilities of such vision-language models (VLMs). Popular multimodal models often fail at surprisingly simple visual tasks where humans achieve near-perfect performance. Research shows that VLMs struggle even with basic spatial relations such as deciding what is left or right in an image, with performance close to random guessing on some benchmarks. In this work, we take a deep dive into investigating the fundamental causes of such shortcomings from a vision perspective. In particular, we study the most common image encoders leveraged by VLMs and propose a technique to "see through the eyes" of the model, which we call CLIP-Ocular. This allows us to gain new insights into VLM shortcomings in a principled manner. We then utilize these new insights to redesign the VLM training pipeline to enhance their spatial reasoning capabilities.

3 - Maximum Entropy Solution of Linear Systems

Michael Friedlander, University of British Columbia, Vancouver, BC, Canada

A Newton-type method is described for obtaining a maximum-entropy solution over the probability simplex to an underdetermined linear system. The method accomodates direct and indirect linear algebra operations, including preconditioning. Numerical experiments on synthetic and real data confirm the accuracy and scalability of the approach.

Joint work with Nicholas Barnfield and Tim Hoheisel.

4 - Sum-of-Minimum Optimization

Wotao Yin, DAMO Academy, Alibaba US, Bellevue, WA, United States, Lisang Ding, Ziang Chen, Xinshang Wang

Machine learning (ML) typically trains a single model by minimize the average training loss across all data points. This approach is sub-optimal when dealing with heterogeneous data or when the data come from multiple sources. We propose a novel "sum-of-minimum" optimization model by considering an ensemble of k models and optimizing them jointly. The key idea is to find the optimal assignment of each data point to the model that performs best on it, while simultaneously training these models. This approach is mathematically formulated as:

minimize $\text{sum}_{i=1,..,n} \min\{\text{loss}(x_1, \text{data}_i), ..., \text{loss}(x_k, \text{data}_i)\}$

where $x_1,...,x_k$ are decision variables that represent the parameters of the k models. By minimizing the sum of minimum losses, each of the k model is incentivized to specialize in a subset of the data points where it achieves the lowest loss.

Although this optimization problem is both non-smooth and non-convex, we present an algorithm that approximately solves the problem with guarantees. Under favorable assumptions, we provide tight performance and convergence bounds for our algorithm.

We demonstrate the effectiveness of our approach through experiments on various tasks, including generalized principal component analysis, multi-neural network training, and mixed linear regression. We believe that the "sum-of-minimum" model can significantly improve ML performance when dealing with diverse and complex datasets.

MB32

Summit - 423

Optimization under Non-Stationary Environments

Invited Session

OPT: Global Optimization

Chair: Dabeen Lee, KAIST, Daejeon

1 - Reinforcement Learning for Infinite-Horizon Average-Reward MDPs with Multinomial Logistic Function Approximation

JAEHYUN PARK, KAIST, Daejeon, Korea, Republic of

We study model-based reinforcement learning with non-linear function approximation where the transition function of the underlying Markov decision process (MDP) is given by a multinomial logistic (MNL) model. In this paper, we develop two algorithms for the infinite-horizon average reward setting. Our first algorithm **UCRL2-MNL** applies to the class of communicating MDPs and achieves an **O(dD√ T)** regret, where **d** is the dimension of feature mapping, **D** is the diameter of the underlying MDP, and **T** is the horizon. The second algorithm **OVIFH-MNL** is computationally more efficient and applies to the more general class of weakly communicating MDPs, for which we show a regret guarantee of **O(d 2/5 sp(v*)T 4/5)** where **sp(v*)** is the span of the associated optimal bias function. We also prove a lower bound of **Ω(d √ DT)** for learning communicating MDPs with MNL transitions of diameter at most **D**. Furthermore, we show a regret lower bound of **Ω(dH3/2√K)** for learning **H**-horizon episodic MDPs with MNL function approximation where **K** is the number of episodes, which improves upon the best-known lower bound for the finite-horizon setting.

2 - Parameter-Free Algorithms for Performative Regret Minimization under Decision-Dependent Distributions

Sungwoo Park, KAIST, Daejeon, Korea, Republic of

TBD

3 - Nonconvex Stochastic Optimization under Decision-Dependent Distributions

Junyeop Kwon, KAIST, Daejeon, Korea, Republic of

TBD

4 - Stochastic-Constrained Stochastic Optimization with Markovian Data

Yeongjong Kim, POSTECH, Pohang, Korea, Republic of, Dabeen Lee

This paper considers stochastic-constrained stochastic optimization where the stochastic constraint is to satisfy that the expectation of a random function is below a certain threshold. In particular, we study the setting where data samples are drawn from a Markov chain and thus are not independent and identically distributed. We generalize the drift-plus-penalty framework, a primal-dual stochastic gradient method developed for the i.i.d. case, to the Markov chain sampling setting. We propose two variants of drift-plus-penalty; one is for the case when the mixing time of the underlying Markov chain is known while the other is for the case of unknown mixing time. In fact, our algorithms apply to a more general setting of constrained online convex optimization where the sequence of constraint functions follows a Markov chain. Both algorithms are adaptive in that the first works without knowledge of the time horizon while the second uses AdaGrad-style algorithm parameters, which is of independent interest. We demonstrate the effectiveness of our proposed methods through numerical experiments on classification with fairness constraints.

MB33

Summit - 424

Data-Driven Methods for Robust and Distributionally Robust Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Irina Wang, Princeton University, Princeton, NJ, United States

Co-Chair: Bartolomeo Stellato, Princeton University, Princeton, NJ, United States

1 - Beyond Absolute Continuity: a New Class of Dynamic Risk Measures

Jincheng Yang, The University of Chicago, Chicago, IL, United States, Rui Gao, Luhao Zhang

The modern theory of risk measures copes with uncertainty by considering multiple probability measures. While it is often assumed that a reference probability measure exists, under which all relevant probability measures are absolutely continuous, there are examples where this assumption does not hold, such as certain distributional robust functionals. In this talk, we introduce a novel class of dynamic risk measures that do not rely on this assumption. We will discuss its convexity, coherence, and time consistency properties.

2 - A Decision Rule Approach for Two-Stage Data-Driven Distributionally Robust Optimization Problems with Random Recourse

Xiangyi Fan, ExxonMobil, Spring, TX, United States, Grani Adiwena Hanasusanto

This research investigates two-stage stochastic optimization problems with random recourse, where the adaptive decisions are multiplied by the uncertain parameters in both the objective function and the constraints. To mitigate the computational intractability of infinite-dimensional optimization, we propose a scalable approximation scheme via piecewise linear and piecewise quadratic decision rules. Based on the decision rule structure, we develop a data-driven distributionally robust framework with two layers of robustness to address distributional uncertainty. Additionally, we establish out-of-sample performance guarantees for our proposed scheme. The resulting optimization problem can be reformulated as an exact copositive program, which allows for tractable approximations using semidefinite programming. Furthermore, we design a decomposition algorithm that enables parallel solution of smaller-size semidefinite programs, leading to reduced runtime. Through numerical examples, we empirically demonstrate that our method produces significantly better solutions with reasonable computational effort than the traditional sample-average approximation scheme, especially when the data is limited.

3 - Learning Decision-Focused Uncertainty Sets in Contextual Dynamic Optimization

Irina Wang, Princeton University, Princeton, NJ, United States, Bartolomeo Stellato

We propose a data-driven technique to automatically learn the uncertainty sets in robust optimization. Specifically, the uncertainty sets learned are based on contextual information and for a parametric family of adaptive robust optimization problems; using machine learning, we learn a mapping from context parameters to context-specific uncertainty sets. We formulate the problem using stochastic bi-level optimization, where in the outer level we minimize the expected loss of the data-driven decisions across the parametric family, subject to guaranteeing constraint satisfaction. In the lower level, we solve the parametric robust optimization problems for the data-driven decisions. Our approach is very flexible and can learn a wide variety of uncertainty sets while preserving tractability. We solve the constrained learning problem using a stochastic augmented Lagrangian method, and show convergence, under mild conditions, to a feasible solution. Numerical experiments show that our method outperforms traditional and non-contextual approaches in robust optimization in terms of out-of-sample performance and constraint satisfaction guarantees.

4 - Neur2RO: Neural Two-Stage Robust Optimization

Justin Dumouchelle, University of Toronto, Toronto, ON, Canada, Esther Julien, Jannis Kurtz, Elias B. Khalil

Robust optimization provides a mathematical framework for modeling and solving decision-making problems under worst-case uncertainty. This work addresses two-stage robust optimization (2RO) problems (also called adjustable robust optimization), wherein first-stage and second-stage decisions are made before and after uncertainty is realized, respectively. This results in a nested min-max-min optimization problem which is extremely challenging computationally, especially when the decisions are discrete. We propose Neur2RO, an efficient machine learning-driven instantiation of column-and-constraint generation (CCG), a classical iterative algorithm for 2RO. Specifically, we learn to estimate the value function of the second-stage problem via a novel neural network architecture that is easy to optimize over by design. Embedding our neural network into CCG yields high-quality solutions quickly as evidenced by experiments on two 2RO benchmarks, knapsack and capital budgeting. For knapsack, Neur2RO finds solutions that are within roughly 2% of the best-known values in a few seconds compared to the three hours of the state-of-the-art exact branch-and-price algorithm; for larger and more complex instances, Neur2RO finds even better solutions. For capital budgeting, Neur2RO outperforms three variants of the k-adaptability algorithm, particularly on the largest instances, with a 10 to 100-fold reduction in solution time. In addition, extensions to single-stage robust optimization will be presented.

5 - Enhancing Scalability for Two-Stage Distributionally Robust Optimization over 1-Wasserstein Balls

Geunyeong Byeon, Arizona State University, Tempe, AZ, United States

We present a decomposition algorithm along with several algorithmic enhancements designed to address two-stage distributionally robust optimization problems over 1-Wasserstein balls. Through a numerical analysis on a facility location problem with demand uncertainty, we highlight the potential benefits of employing a particular distance metric for enhanced out-of-sample performance, while also examining its computational trade-offs. Furthermore, we explore alternative approaches aimed at enhancing the scalability of our proposed algorithm.

MB34

Summit - 425

Optimization Issues in Recent AI Models

Invited Session

OPT: Nonlinear Optimization

Chair: Yushun Zhang, The Chinese University of Hong Kong, Shenzhen, shenzhen, 518000

Co-Chair: Jiawei Zhang, MIT, Cambridge, MA, United States

1 - Hardware-Efficient Deep Learning Algorithms: Foundation and Practice

Jiawei Zhao, Meta FAIR, Pasadena, CA, United States

In recent years, neural networks have demonstrated impressive performance across various disciplines, including conversational AI and scientific discovery. Training has become increasingly challenging due to the growing size of parameters and computational demands. On the other hand, the optimization algorithms (e.g., SGD and Adam), which are fundamental to deep learning, have remained largely unchanged. Although powerful, these classical algorithms may be inadequate for addressing the recent challenges in large-scale training, where hardware limitations are a primary bottleneck. In this talk, I will discuss my work on developing hardware-efficient learning algorithms that jointly consider the optimization and hardware constraints. Specifically, I will introduce my recent work GaLore, for memory-efficient training of large language models. GaLore projects gradients into low-rank subspaces to reduce memory requirements for optimizer states and gradients during training. It reduces total training memory by 63.3% while preserving modeling performance, as demonstrated in training LLaMA models from scratch on the C4 dataset. GaLore makes it feasible to pre-train the LLaMA 7B model on consumer GPUs with 24GB of memory (e.g., NVIDIA RTX 4090) without the need for model parallelism, checkpointing, or offloading strategies. In closing, I will discuss the challenges and opportunities for further enhancing the efficiency of training large language models.

2 - Analyzing the deep learning optimization in the easy way

Yingyu Liang, The University of Hong Kong, Hong Kong, China, People's Republic of

Deep learning is a key factor in the recent breakthroughs in artificial intelligence. Despite the empirical success, the theoretical understanding remains largely elusive, especially because of the challenges in analyzing the optimization involved. This talk will discuss why optimization is key for understanding deep learning, review the NTK analysis approach and its limitations, and present our recent effort to go beyond that.

3 - Landscape evolution of neural networks through the lens of neuron splitting

Tian Ding, Shenzhen Institute of Big Data, Shenzhen, China, People's Republic of

In the field of artificial intelligence, the size of neural networks has rapidly increased, resulting in deeper and wider models. While it is believed that increasing the width of a neural network enhances its training efficiency, the underlying mechanisms remain not fully understood. In this presentation, we investigate the evolution of the optimization landscape as network width increases. Specifically, we establish relationships between stationary points in shallow and wide networks through an operator known as "neuron splitting." Our analysis yields three theoretical insights as neural networks become wider. First, stationary points at the same loss levels become interconnected. Second, local minima transform into plateaus that predominantly include saddle points but may also contain local minima. Third, when the network width exceeds a certain threshold, there exists a non-increasing path from any stationary point to a global minimum. These findings provide a new understanding of the behavior and potential of wide neural networks, contributing to the ongoing development of more effective AI models.

4 - Diffusion Model for Data-Driven Black-Box Optimization

Zihao Li, Princeton University, Princeton, NJ, United States

Generative AI has redefined artificial intelligence, enabling the creation of innovative content and customized solutions that drive business practices into a new era of efficiency and creativity. In this paper, we focus on diffusion models, a powerful generative AI technology, and investigate their potential for black-box optimization over complex structured variables. Consider the practical scenario where one wants to optimize some structured design in a high-dimensional space, based on massive unlabeled data (representing design variables) and a small labeled dataset. We study two practical types of labels: 1) noisy measurements of a real-valued reward function and 2) human preference based on pairwise comparisons. The goal is to generate new designs that are near-optimal and preserve the designed latent structures. Our proposed method reformulates the design optimization problem into a conditional sampling problem, which allows us to leverage the power of diffusion models for modeling complex distributions. In particular, we propose a reward-directed conditional diffusion model, to be trained on the mixed data, for sampling a near-optimal solution conditioned on high predicted rewards. Theoretically, we establish sub-optimality error bounds for the generated designs. The sub-optimality gap nearly matches the optimal guarantee in off-policy bandits, demonstrating the efficiency of reward-directed diffusion models for black-box optimization. Moreover, when the data admits a low-dimensional latent subspace structure, our model efficiently generates high-fidelity designs that closely respect the latent structure. We provide empirical experiments validating our model in decision-making and content-creation tasks.

5 - Tuning-Free Stochastic Optimization

Ahmed Khaled, Princeton University, Princeton, NJ, United States, Chi Jin

Hyperparameter tuning is too costly to use in today's large-scale machine learning problems. This creates a need for algorithms that can tune themselves on-the-fly. We formalize the notion of "tuning-free" algorithms that can match the performance of optimally-tuned optimization algorithms up to polylogarithmic factors given only loose hints on the relevant problem parameters. We consider in particular algorithms that can match optimally-tuned Stochastic Gradient Descent (SGD). When the domain of optimization is bounded, we show tuning-free matching of SGD is possible and achieved by several existing algorithms. We prove that for the task of minimizing a convex and smooth or Lipschitz function over an unbounded domain, tuning-free optimization is impossible. We discuss conditions under which tuning-free optimization is possible even over unbounded domains. In particular, we show that the recently proposed DoG and DoWG algorithms are tuning-free when the noise distribution is sufficiently well-behaved. For the task of finding a stationary point of a smooth and potentially nonconvex function, we give a variant of SGD that matches the best-known high-probability convergence rate for tuned SGD at only an additional polylogarithmic cost. However, we also give an impossibility result that shows no algorithm can hope to match the optimal expected convergence rate for tuned SGD with high probability.

MB35

Summit - 427

Panel: Practitioners' View on Transportation and Logistics Research

Panel Session

Transportation Science and Logistics (TSL)

Co-Chair: Sibel Alumur Alev, University of Waterloo, Waterloo, ON, Canada

1 - Panelist

Tim Jacobs, Amazon, Tempe, AZ, United States

2 - Panelist

Philip Kaminsky, Amazon, El Cerrito, CA, United States

3 - Panelist

Anne Robinson, Robinson Insights, Ottawa, ON, Canada

4 - Panelist

Ruben Lobel, Waymo, San Francisco, CA, United States

MB36

Summit - 428

Facility Design and Operations in Emerging Mobility Services

Invited Session

TSL: Facility Logistics

Chair: Xinyu Liu, Georgia Tech ISyE, Atlanta, GA, United States

1 - Daily Opportunity Charging Scheduling for Electric Buses: Exact and Heuristic Methods

Dan McCabe, University of Washington, Seattle, WA, United States, Xuegang Ban, Balazs Kulcsar

Transit agencies across the world are transitioning to fully electric bus fleets. For many of these agencies, current battery range is not sufficient to cover all daily driving needs, so high-power opportunity charging will be an essential tool to extend effective range. However, introducing this extra activity into the service day can negatively impact passenger service if it is not planned carefully. Because opportunity chargers are expensive and scarce, uncoordinated charging can lead to extensive queuing at charging stations, creating delays for passenger trips later in the day. To avoid these issues, transit system operators should plan charging schedules in advance each day that minimize passenger delays while ensuring all buses stay sufficiently charged.

This work generates opportunity charging schedules with a novel model that directly tracks queuing at chargers in order to set and propagate departure delays. We solve the complex MILP model with two tailored algorithms (one exact, one heuristic) based on decomposition. Computational tests on both a simple notional transit network and the real-world bus system of King County, Washington demonstrate our methods' effectiveness. The heuristic algorithm in particular, which runs in polynomial time, is shown to generate near-optimal solutions quickly and scales well enough to apply to large bus networks.

2 - Operational Policy Design and Evaluation for Connected Pudo Facilities

Xinyu Liu, Georgia Tech ISyE, Atlanta, GA, United States

Pickup and drop-off (PUDO) facilities have long been essential in facilitating passenger transportation and goods delivery. Emerging mobility services, such as on-demand hailed rides and time-sensitive deliveries, create a need for frequent pickups and drop-offs of passengers and small shipments close to the end consumers. The pickups and drop-offs of passengers and goods are frequently observed at limited designated spaces on the curb. The scarcity of such spaces in dense urban areas, as opposed to the increasing demand, leads to safety and efficiency concerns arising from double-parking behaviors and unnecessary vehicle detours. Thus, carefully designed and operated PUDO facilities are much needed to fulfill this emerging demand. In this talk, I will introduce mathematical and computational models to optimize the long-term performance metrics under uncertainty. Unlike traditional quantitative methods, our stochastic models capture realistic vehicle conflicts and dependencies, thus unveiling new insights into facility operations. Our work also proposes and evaluates practical, interpretable, and implementable approximate operational policies with demonstrated near-optimality, envisioning a future with enhanced connectivity and computational power.

3 - Optimizing Autonomous Transfer Hub Networks: Quantifying the Potential Impact of Self-Driving Trucks

Kevin Dalmeijer, Georgia Institute of Technology, Atlanta, GA, United States, Chungjae Lee, Pascal Van Hentenryck, Peibo Zhang

Autonomous trucks are expected to fundamentally transform the freight transportation industry. In particular, Autonomous Transfer Hub Networks (ATHNs), which combine autonomous trucks on middle miles with human-driven trucks on the first and last miles, are seen as the most likely deployment pathway for this technology. This paper presents a framework to optimize ATHN operations and evaluate the benefits of autonomous trucking. By exploiting the problem structure, this paper introduces a flow-based optimization model for this purpose that can be solved by blackbox solvers in a matter of hours. The resulting framework is easy to apply and enables the data-driven analysis of large-scale systems. The power of this approach is demonstrated on a system that spans all of the United States over a four-week horizon. The case study quantifies the potential impact of autonomous trucking and shows that ATHNs can have significant benefits over traditional transportation networks.

4 - Noise-Aware and Equitable Urban Air Traffic Management: An Optimization Approach

Zhenyu Gao, The Hong Kong University of Science and Technology, Hong Kong SAR, Hong Kong, Yue Yu, Qinshuang Wei, Ufuk Topcu, John-Paul Clarke

Urban air mobility (UAM), a transformative concept for the transport of passengers and cargo, faces several integration challenges in complex urban environments. Community acceptance of aircraft noise is among the most noticeable of these challenges when launching or scaling up a UAM system. Properly managing community noise is fundamental to establishing a UAM system that is environmentally and socially sustainable. In this work, we develop a holistic and equitable approach to manage UAM air traffic and its community noise impact in urban environments. The proposed approach is a hybrid approach that considers a mix of different noise mitigation strategies, including limiting the number of operations, cruising at higher altitudes, and ambient noise masking. We tackle the problem through the lens of network system control and formulate a multi-objective optimization model for managing traffic flow in a multi-layer UAM network while concurrently pursuing demand fulfillment, noise control, and energy saving. Further, we use a social welfare function in the optimization model as the basis for the efficiency-fairness trade-off in both demand fulfillment and noise control. We apply the proposed approach to a comprehensive case study in the city of Austin and perform design trade-offs through both visual and quantitative analyses.

MB37

Summit - 429

Simulation and Learning for Smart Transportation Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Li Jin, 800 Dongchuan Lu, 200240

Co-Chair: Shenhao Wang, University of Florida, Gainesville, FL, United States

1 - Combining AI and Network Science for Transportation Network Planning

Zhan Zhao, University of Hong Kong, Hong Kong, China, People's Republic of

The advancement in machine learning and artificial intelligence (AI) technologies, coupled with the increasing prevalence of mobility big data, has profoundly transformed the ways we analyze, and understand, our transport systems. While current AI applications in transportation mostly focus on short-term demand forecasting and operations, long-term demand forecasting and transportation planning tasks have received less attention. As transportation systems are often organized in networks, it is necessary to account for the intricate spatial interactions along the network space in transportation planning. In this talk, I will discuss how AI can be combined with network science to enhance long-term travel demand forecasting to support transportation network planning decisions. Specifically, I will introduce a series of case studies, including the planning of bike sharing systems, metro systems, and street networks.

2 - Modeling Urban Traffic Data with Matrix and Tensor Factorization Approaches

Xinyu Chen, Massachusetts Institute of Technology, Cambridge, MA, United States

In urban systems, large amounts of human movement data, such as urban mobility and traffic flow, are readily available for implementing downstream tasks and supporting decision-making. Since these data are characterized by spatiotemporal dimensions and demonstrate data patterns of the systems, it is important to reformulate urban system problems with machine learning. In this talk, we introduce two scentific problems: 1) imputing missing values from partially observed traffic data, and 2) discovering spatial and temporal patterns from time-varying urban systems. Since these spatiotemporal data always (empirically) reveal low-rank properties and spatiotemporal correlations, we start from a sequence of low-rank matrix and tensor methods such as matrix/tensor factorization. To reinforce the model for building temporal correlations and dynamics, we integrate time series convolution and autoregression into the framework. In the modeling process, we intend to highlight the importance of temporal modeling in the low-rank matrix and tensor methods.

3 - Machine Learning for Large-Scale Combinatorial Optimization

Sirui Li, Cambridge, MIT, MA, United States

Combinatorial Optimization (CO) is widely used in transportation, with applications in Vehicle Routing, Scheduling to Logistics and Production Planning. However, solving CO problems can be time consuming. For instance, it takes over one hour to solve a Vehicle Routing problem with 2000 locations. In this talk, I will present several of my PhD projects on using machine learning to speed up combinatorial optimization. These include: (1) Learning to Delegate for Large-scale Vehicle Routing. This is a novel learning-augmented local search framework designed to solve large-scale VRPs. The method iteratively improves the solution by identifying appropriate subproblems and delegating their improvement to a black box subsolver. Our method accelerates state-of-the-art VRP solvers by 10x to 100x while achieving competitive solution qualities for VRPs with sizes ranging from 500 to 3000. (2) Learning to Configure Separators in Branch-and-Cut. Here, we focus on general-purpose Mixed Integer Linear Programming (MILP) solvers. We identify that MILP solvers can be drastically accelerated by appropriately selecting separators to activate. As the combinatorial separator selection space imposes challenges for machine learning, we learn to separate by proposing a novel data-driven strategy to restrict the selection space and a learning-guided algorithm on the restricted space. Our method predicts instance-aware separator configurations which can dynamically adapt during the solve, effectively accelerating the open source MILP solver SCIP by improving the relative solve time up to 72% and 37% on synthetic and real-world MILP benchmarks.

4 - Towards Financial Viability for Electric Mobility: Explore Profitability Models for Electric Vehicle Owners

Anqi Han, Hong Kong University of Science and Techonology(Guangzhou), Guangzhou, China, People's Republic of

Efforts to mitigate climate change enable Electric Vehicle (EV) as a sustainable alternative of transportation to conventional ones. Meanwhile, Vehicle-to-Grid (V2G) technology provides EV stakeholders with an energy trading market, where EVs can provide services directly to the grid. Sets of business models have been under discussion to create economic values for V2G integration. To bridge the gap between theoretical V2G concepts and practical applications, this paper chose to study Vehicle-to-Building (V2B) configurations, extending to private and public charging scenarios. We emphasize exploring EV owner's profitability under arbitrage trading conditions, where EVs are charged either at home or at public charging stations, and subsequently provide energy to workplace. First, a synthetic dataset of EV travel profiles is generated using the Monte Carlo simulation approach, incorporating stochastic elements to reflect real-world variability. Then actions to charge, discharge, or to take no actions are determined by EV's State of Charge (SoC) and its daily trip constrains. Next, a mixed integer Linear Programming approach is applied to model the profitability of EV owners, considering the total costs embedded in each charge/discharge cycle and cumulative returns gained from V2G participation. Specifically, the novel contribution of this paper lies in attaching financial attributes to EVs. The core idea is to design a progressive time-based V2G incentive pricing that reflects the value of a vehicle's availability time to discharge. Our findings indicate that incorporating financial attributes into EVs not only underscores the economic viability for individual EV owners but also facilitates the practical integration of V2G technologies.

5 - Optimizing Food Delivery Services: Reinforcement Learning for Enhanced Efficiency and Equity

Yi Fan, Tongji University, Shanghai, China, People's Republic of

Food delivery services have become very popular in recent years. As delivery demand increases exponentially, optimizing the efficiency and safety of these services has become increasingly crucial. In this work, we utilize a combination of operations research and artificial intelligence within a gridded rider delivery environment to analyze and predict rider delivery route selection and order preferences. By integrating these factors into the objective function of our reinforcement learning model, we assess the reasonableness of the current dispatch data used by Meituan. Ultimately, by explicitly accounting for rider behavior and transportation equity, the platform can develop and optimize dispatch assignments that lower costs for the system as a whole – a rare win-win outcome.

MB38

Summit - 430

Machine Learning Application in Aviation

Invited Session

Aviation Applications

Chair: Xiyitao Zhu, University of Illinois Urbana-Champaign, Champaign, IL, 61822, United States

1 - Learning for Aircraft Trajectory Optimization in the Airspace

Xiyitao Zhu, University of Illinois Urbana-Champaign, Champaign, IL, United States, Jing Gao, Lavanya Marla, Ankur Mani

In the US National Aviation System (NAS), aircraft rely on weather information for routing decisions. We propose that using en-route aircraft as sensors to collect updated weather information at the right place and right time maintains high-quality weather information that benefits the NAS. Motivated by the spatial and temporal correlations in weather, we propose a novel modeling framework for routing decisions on (airspace) networks with non-stationary Gaussian bandits, temporal information diffusion, and in particular, with combinatorial path structures. Our framework is also applicable to intra-city aircraft and unmanned aircraft such as UAVs and drones.

2 - ML models for hotspots' identification

Guglielmo Lulli, University of Milano Bicocca, Milano, Italy, Sameer Alam, Davide Falanga, Giacomo Lancia, Bizhao Pang

In this talk, I will present some of the outcomes of the Italy-Singapore AI4ATM research project, co-funded by MIUR and A*STAR, focusing on the application of machine learning models for the analysis of air traffic trajectories and hotspots' identification and prediction. Additionally, I will discuss strategies for resolving these hotspots.

3 - Benefits of Slot Substitutions for Airlines during Ground Delay Programs

Jing Xu, University of California, Berkeley, Berkeley, CA, United States

Slot substitution is one part of Collaborate Decision Making (CDM). This policy provides airlines some flexibility during Ground Delay Programs (GDPs). This paper simulates how airlines perform slot substitutions and estimates benefits of slot substitutions for American Airlines during GDPs. We assume an optimization model to minimize delay cost for airline slot substitution. By synthetic data and substitution simulation, we calibrate the model by finding parameter values that produce a close match between modeled and observed distributions of slot substitution frequencies for American Airlines, Delta Airlines, and United Airlines. We can infer that the three major airlines may willing to sacrifice a maximum of 7.34 hours of delay for each flight to avoid step-costs caused by misconnection issues. These three airlines saves about $49.4 million dollars from being able to perform slot substitutions in 991 GDPs in 2019. The savings from slot substitution are sensitive to GDP airports.

4 - Feature-based Pricing for Airline Ancillary Products

Xiao JIN, NUS Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Changchun Liu, Chung Piaw Teo

This study addresses the challenge of feature-based pricing with limited price experiments. As the airline industry strives for customized pricing through refined feature-based customer segmentation, calibrating the underlying models becomes more challenging. With limited seats per flight, the maximum transactional data acquired remains constrained. Traditional calibration methods, such as Maximum Likelihood Estimation (MLE), often fail to provide stable feature-based pricing under these conditions. Our approach addresses this stability issue by calibrating a feature-based Marginal Exponential Model (MEM) through minimizing market share discrepancy. This method stabilizes with less data, enabling a reliable pricing for each segment. Applied to luggage pricing for Flyscoot, a low-cost carrier, our feature-based pricing strategy assesses how each feature influences customer choices, leading to more profitable pricing decisions. These findings suggest that when data is limited, calibration that minimizes share discrepancy is not only feasible but also superior in terms of solvability and stability. This approach offers a refined solution for airlines aiming to enhance ancillary pricing strategies in a competitive market. Additionally, this method can extend to other industries where feature-based pricing and customer segmentation are essential but constrained by limited price experiments.

MB39

Summit - 431

Railway Technology Innovations and Modeling

Invited Session

Railway Applications

Chair: Carl Van Dyke, TransNetOpt, 6 Snowbird Ct, West Windsor, NJ, 08550, United States

1 - Changing Tank Car Markets and Demand Drivers

Daniel Windle, TrinityRail, Dallas, TX, United States

Compared to the shale boom of the mid-2010s, the market drivers for tank car demand are more varied and will require coordination between shippers, car owners, and railroads to make sure the market is served. This presentation will review some of the drivers for tank car demand and what analytical solutions may be needed to start to account for these changes.

2 - OR-related fleet modeling approaches at TTX Co.

Muhannad Ramahi, TTX Company, Chicago, IL, United States

Automobile manufacturers (shippers) provide railroad companies with annual forecasts of their monthly shipping volumes from various origins to different destinations. The railroad companies (carriers) jointly operate pools of railcars to transport automobiles. Each pool comprises equipment of a particular type and serves one or more shippers. RELOAD, a fleet management group within TTX (an independent company), manages the repositioning of empty railcars of each type for the carriers. The problem is to determine the smallest fleet size that will provide adequate service

3 - Using AI for cleaning of waybills and other data

Daril Vilhena, CEDAR, Seattle, WA, United States

Railroads have significant issues with inconsistent and bad data, as well as applying various overrides to shipping instructions. Traditionally these issues have been addressed by using static rules tables and replacement (alias) tables to "fix" or "override" the offending data fields. AI offers the opportunity to automate this correction process, reducing or eliminating the cumbersome rules tables, improving accuracy, maintainability of their systems and reducing labor requirements.

4 - Parallel Systems' Autonomous Rail Vehicles: Development and Applications

Marty Schlenker, Parallel Systems, Mechanicsburg, PA, United States

Marty Schlenker, Head of Strategy for Parallel Systems, will provide an update on the development of Parallel Systems' autonomous, battery-electric rail vehicles, including Parallel's viewpoints on their ability to attract freight from truck to rail in specific applications and their interoperation with conventional trains on rail networks.

MB40

Summit - 432

Responsible Research in Behavioral Operations

Invited Session

Behavioral Operations Management

Chair: Karen Donohue, University of Minnesota, Minneapolis, MN, United States

Co-Chair: Yeonjoo Lee, University of Minnesota, Minneapolis, MN, United States

1 - Nudging Green-but-Slow Shipping in Online Retail

Yeonjoo Lee, University of Minnesota, Minneapolis, MN, United States, Karen Donohue

Faster e-commerce fulfillment often comes at an environmental cost by using more energy-intensive transportation modes or more frequent, less consolidated, last mile delivery. While faced with these externalities, retailers still feel pressure to provide fast delivery to stay competitive. But, what if consumers could be nudged to choose a green but slow shipping option? In practice, a variety of information strategies are used to nudge such decisions, but it remains unclear which strategy fits a given logistical context and why. We examine these questions in two logistical contexts: no-rush shipping and consolidated shipping. Drawing from practice and behavioral research, we identify a comprehensive set of information strategies for both contexts and develop a theory to inform effective strategies based on their ability to overcome psychological barriers.

2 - Selection, Payment, and Information Assessment in Social Audits: A Behavioral Experiment

Gabriel Alfonso Pensamiento Calderon, University of Pittsburgh, Pittsburgh, PA, United States, Leon Valdes

Companies often rely on third-party social audits to assess suppliers' social responsibility (SR) practices. However, empirical evidence suggests these audits can often be too lenient and poor practices go unreported, particularly if the auditor feels beholden to the supplier. We design and conduct an incentivized lab experiment to study how a supplier choosing and/or paying the auditor affects audit reports. We explore whether and how these levers affect auditors' assessment of noisy signals about the supplier's practices, and investigate the role of two behavioral phenomena: motivated reasoning and reciprocity. We find that auditors who are paid and chosen by the supplier are more lenient, with the effect more pronounced when the information observed suggests poor SR practices. We do not find evidence of auditors being more likely to believe the supplier has good practices when chosen by it. Conversely, reciprocity towards the supplier plays an important role behind our results. Finally, auditors who are merely paid by the supplier are not more lenient, which offers good news for practitioners. Our results can help guide companies' auditing and procurement policies. First, they show that removing a supplier's ability to choose its own auditor is critical to increase detection of poor SR practices, particularly when the risk of them is high. Second, removing the suppliers' ability to choose their auditor—while still paying them—might be enough to reduce leniency. Finally, our findings shed light on some of the behavioral drivers behind auditors decisions, which can guide interventions to mitigate audit leniency.

3 - Using Behavioral Nudges to Reduce Food Wastage in Ghana: Evidence from a Field Experiment

Mahyar Eftekhar, Arizona State University, Tempe, AZ, United States, Xinming Liu, Yanchong Zheng, Richard Boso

Research indicates that food waste is a global issue affecting countries across various income levels, with about 931 million tonnes—or 17% of consumer-available food—being discarded in 2019. Meanwhile, nearly 3 billion people worldwide cannot afford a nutritious diet. This study showcases a straightforward, practical method aimed at reducing food waste by focusing on the food preparation processes in restaurants, which contribute to approximately 26% of the total waste.

4 - Matching Volunteers with Clients in a Non-Profit Organization

Ravi Subramanian, Georgia Tech, Atlanta, GA, United States, Shikha Safaya, Basak Kalkanci

Non-profit organizations (NPOs) are often challenged with volunteer participation and retention in the absence of monetary incentives. Based on rich survey data collected in collaboration with our NPO partner, we observed that volunteers and clients have heterogenous preferences: some prefer the base service only, while others prefer the differentiated service that includes interpersonal interactions besides the basic meal delivery service. We build upon this insight to study volunteer management policies by comparing two task assignment policies: a pooled policy that groups volunteers together irrespective of their task preferences; and a preference-based policy that incorporates volunteer preferences for task assignment. Our study explores the tradeoff between incorporating volunteer preferences in task assignments versus pooling volunteers to alleviate mismatch between supply and demand. We endogenize volunteers' participation decisions based on their utility from volunteering as well as their outside commitments. We analytically derive the conditions under which a particular policy may be preferred by all the stakeholders (the non-profit, volunteers, and clients). Our results shed light on managerial levers that NPOs can deploy to improve client welfare and volunteer utility. Thus, our model insights highlight alternatives to simply pooling volunteers into a single bucket, and offer prescriptive insights for NPOs to advance their service missions and enhance volunteer engagement.

MB41

Summit - 433

Advancements in Optimal Control of Queues: Innovations and Applications

Invited Session

Applied Probability Society

Chair: Hayriye Ayhan, ISyE Georgia Tech, Atlanta, GA, United States

Co-Chair: Xinchang Wang, Washington State University, Pullman, WA, United States

1 - A Randomized Routing Control for A Non-Stationary Finite Horizon Multi-Class and Multi-Server Queueing System

Amy Cochran, University of Wisconsin Madison, Madison, WI, United States, Gabriel Zayas-Caban

We consider a non-stationary finite-horizon problem with multiple job classes and finite-capacity service facilities, where any number of jobs may be assigned at any time, there are facility-dependent constraints on the number of class groups allowed in each facility, and service rates are facility-dependent. A heuristic for routing jobs in this system is proposed and shown to be asymptotically optimal.

2 - Discrete-Review Control of Queueing Networks

Mir Mikdad Talpur, University of Toronto, Toronto, ON, Canada, Vahid Sarhangian

Optimal control of stochastic queueing networks which are used in modeling and analysis of service systems (e.g., hospitals, contact centers, and ride/bike sharing systems) has been the subject of extensive research. The literature is predominantly concerned with continuous-time control, whereby the controller can interact with the system at any point in time. In many service settings, however, the system manager can only interact with the system at discrete and possibly long time-intervals. That is, although relevant information is becoming available, the controller cannot exercise control before the beginning of the next review time. Our main objective is to develop fundamental insights on the impact of imposing the discrete-review limitation on the performance of dynamic control policies. A key question we will answer is the effect on the value function of the optimal control problem as the length of the review period is varied, for a general formulation of queueing control problems. We work with a generic continuous-time control problem where we restrict operational control to the start of fixed time-intervals. Due to the intractability of the transition probabilities under discrete-time control in an exact Markov Decision Processes (MDP) approach that makes the analysis intractable, we instead work with the associated fluid control problems that are justified through Functional Strong Law of Large Numbers (FSLLN) for appropriately scaled sequence of the underlying processes. We then investigate partial derivatives of the value function with respect to the period lengths.

3 - Estimating Ship Arrival Rates and Charterer Taste Coefficients with Unobservable Balking

weikun xu, Washington State University, Pullman, WA, United States, Xinchang Wang

This work investigates the joint estimation of the ship arrival rate, service rate, and charterer taste coefficients for a port terminal using ship data. The service process at the terminal, including ship berthing and cargo loading/unloading, is modeled as an M/M/1 queue. Incoming ships decide whether to use the terminal or balk based on a multinomial logit (MNL) model that maps the taste coefficients of ship charterers to choice probabilities. The problem is challenging since the data are censored as balking ships are not observed. We formulate the estimation problem as a maximum likelihood estimation model. We characterize the conditions for model parameters to be identified or uniquely determined. Algorithms are developed to solve the problem and their performances are evaluated against benchmark algorithms.

4 - Optimal Pricing with Impatient Customers

Jieqi Di, ISyE Georgia Tech, Atlanta, GA, United States, Sigrun Andradottir, Hayriye Ayhan

We investigate the optimal pricing strategy in a service-providing framework, where customers can become impatient and leave the system prior to completing service. In this setting, a price is quoted to an incoming customer based on the current number of customers in the system. When the quoted price is lower than the price the incoming customer is willing to pay (which follows a fixed probability distribution), then the customer joins the system and a reward equal to the quoted price is earned. A fixed cost is incurred upon abandonment and a holding cost is incurred for customers waiting to be served. Our goal is to determine the optimal pricing policy that maximizes the long-run average revenue. Our analysis shows that under reasonable assumptions, the optimal quoted price increases as the number of customers in the system grows. Using this insight, we propose a monotone policy iteration algorithm that efficiently computes the optimal pricing policy. Moreover, we introduce a heuristic that charges a fixed price until the number of customers in the system reaches a certain threshold. By selecting the optimal threshold that maximizes the long-run average gain, the heuristic achieves near optimality. We provide an efficient algorithm to find that optimal threshold.

MB42

Summit - 434

The Power of Optimization in Sequential Decision Making

Invited Session

Applied Probability Society

Chair: Adam Block, MIT, Cambridge, MA, United States

Co-Chair: Abhishek Shetty, University of California, Berkeley, Berkeley, CA, United States

Co-Chair: Sasha Rakhlin, MIT, Cambridge, MA, United States

1 - Empirical Risk Minimization Beyond Independent Data

Adam Block, MIT, Cambridge, MA, United States, Sasha Rakhlin, Abhishek Shetty

In order to circumvent statistical and computational hardness results in sequential decision-making, recent work has considered smoothed online learning, where the distribution of data at each time is assumed to have bounded likeliehood ratio with respect to a base measure when conditioned on the history. While previous works have demonstrated the benefits of smoothness, they have either assumed that the base measure is known to the learner or have presented computationally inefficient algorithms applying only in special cases. This work investigates the more general setting where the base measure is \emph{unknown} to the learner, focusing in particular on the performance of Empirical Risk Minimization (ERM) with square loss when the data are well-specified and smooth. We show that in this setting, ERM is able to achieve sublinear error whenever a class is learnable with iid data; in particular, ERM achieves error scaling as
with the statistical complexity of learning the function class with iid data. In so doing, we prove a novel norm comparison bound for smoothed data that comprises the first sharp norm comparison for dependent data applying to arbitrary, nonlinear function classes. We complement these results with a lower bound indicating that our analysis of ERM is essentially tight, establishing a separation in the performance of ERM between smoothed and iid data.

2 - Is Behavior Cloning All You Need? Understanding Horizon in Imitation Learning

Dylan Foster, Microsoft Research, New York, NY, United States, Adam Block, Dipendra Misra

Imitation learning (IL) aims to mimic the behavior of an expert in a sequential decision making task by learning from demonstrations, and has been widely applied to robotics, autonomous driving, and autoregressive language generation. The simplest approach to IL, behavior cloning (BC), is thought to incur sample complexity with unfavorable quadratic dependence on the problem horizon, motivating a variety of different online algorithms that attain improved linear horizon dependence under stronger assumptions on the data and the learner's access to the expert.

In this talk, we revisit the apparent gap between offline and online IL from a learning-theoretic perspective, with a focus on general policy classes up to and including deep neural networks. Through a new analysis of behavior cloning with the logarithmic loss, we will show that it is possible to achieve *horizon-independent* sample complexity in offline IL whenever (i) the range of the cumulative payoffs is controlled, and (ii) an appropriate notion of supervised learning complexity for the policy class is controlled. When specialized to stationary policies, this implies that the gap between offline and online IL is not fundamental. We will then discuss implications of this result and investigate the extent to which it bears out empirically.

3 - Online Estimation via Offline Estimation: An Information-Theoretic Framework

Jian Qian, MIT, Cambridge, MA, United States, Dylan Foster, Yanjun Han, Alexander Rakhlin

The classical theory of statistical estimation aims to estimate a parameter of interest under data generated from a fixed design ("offline estimation"), while the contemporary theory of online learning provides algorithms for estimation under adaptively chosen covariates ("online estimation"). Motivated by connections between estimation and interactive decision making, we ask: is it possible to convert offline estimation algorithms into online estimation algorithms in a black-box fashion? We investigate this question from an information-theoretic perspective by introducing a new framework, Oracle-Efficient Online Estimation (OEOE), where the learner can only interact with the data stream indirectly through a sequence of offline estimators produced by a black-box algorithm operating on the stream. Our main results settle the statistical and computational complexity of online estimation in this framework. Statistical complexity. We show that information-theoretically, there exist algorithms that achieve near-optimal online estimation error via black-box offline estimation oracles, and give a nearly-tight characterization for minimax rates in the OEOE framework. Computational complexity. We show that the guarantees above cannot be achieved in a computationally efficient fashion in general, but give a refined characterization for the special case of conditional density estimation: computationally efficient online estimation via black-box offline estimation is possible whenever it is possible via unrestricted algorithms. Finally, we apply our results to give offline oracle-efficient algorithms for interactive decision making.

4 - Oracle-Efficient Differentially Private Learning with Public Data

Zhiwei Steven Wu, CMU, Pittsburgh, PA, United States

Due to statistical lower bounds on the learnability of many function classes under privacy constraints, there has been recent interest in leveraging public data to improve the performance of private learning algorithms. In this model, algorithms must always guarantee differential privacy for the private samples while also ensuring learning guarantees when the private data distribution is sufficiently close to that of the public data. Previous work has demonstrated that when sufficient public, unlabelled data is available, private learning can be made statistically tractable, but the resulting algorithms have all been computationally inefficient. In this work, we present the first computationally efficient, algorithms to provably leverage public data to learn privately whenever a function class is learnable non-privately, where our notion of computational efficiency is with respect to the number of calls to an optimization oracle for the function class. In addition to this general result, we provide specialized algorithms with improved sample complexities in special cases when the function class is convex or when the task is binary classification.

Joint work with Adam Block, Mark Bun, Rathin Desai, and Abhishek Shetty.

MB43

Summit - 435

Stochastic Systems

Invited Session

Applied Probability Society

Chair: Mark Squillante, IBM Research, Mathematical Sciences, Thomas J. Watson Research Center, Yorktown Heights, NY, 10958, United States

1 - An Efficient High-dimensional Gradient Estimator for Stochastic Differential Equations

Jose Blanchet, Stanford University, Stanford, CA., CA, United States, Shengbo Wang, Peter Glynn

Overparameterized stochastic differential equation (SDE) models have achieved remarkable success in various complex environments, such as PDE-constrained optimization, stochastic control and reinforcement learning, financial engineering, and neural SDEs. These models often feature system evolution coefficients that are parameterized by a high-dimensional vector in the n-dimensional Euclidian space, aiming to optimize expectations of the SDE, such as a value function, through stochastic gradient ascent. Designing efficient gradient estimators for which the computational complexity scales well with n is of significant interest. We introduce a novel unbiased stochastic gradient estimator—the generator gradient estimator—for which the computation time is relatively insensitive to the size n of the parameter space. We discuss unbiasedness and finite variance of the estimator, but also perform numerical experiments that test our estimator in large linear-quadratic control problems parameterized by high-dimensional neural networks. The results show a significant improvement in efficiency compared to the widely used path-wise differentiation method: Our estimator achieves near-constant computation times, increasingly outperforms its counterpart as n increases, and does so without compromising estimation variance.

2 - Long-Time Behavior of Interacting Diffusions on Sparse Graphs

Kevin Hu, Brown University, Providence, RI, United States

We study the long-time behavior of a Conditional McKean-Vlasov (CMVE) equation arising in the study of interacting diffusions on regular trees. We construct a Lyapunov function on the space of probability measures for the weak solution of the CMVE and prove exponential convergence to equilibrium.

3 - Prediction-Enhanced Monte Carlo: A Machine Learning View on Control Variate

Henry Lam, Columbia University, New York, NY, United States, Agostino Capponi, Haoxian Chen, Fengpei Li, Jiahe Lin, Yuriy Nevmyvaka

Despite being an essential tool across engineering and finance, Monte Carlo simulation can be computationally intensive, especially in large-scale, path-dependent problems that hinder straightforward parallelization. A natural attempt is to replace simulation with machine learning or surrogate prediction, but at the expense of a more opaque understanding of the incurred errors. In this talk, we present a Prediction-Enhanced Monte Carlo (PEMC) framework where we leverage machine learning prediction as control variates, thus maintaining unbiased evaluations compared to the direct use of these predictors. On the other hand, unlike existing control variate approaches that rely on mean knowledge and aim for per-sample variance reduction, PEMC takes an expanded perspective on overall-cost-based variance reduction that removes the need of mean knowledge. We demonstrate how PEMC can outperform well-established control variates in traditional problems, and apply in cases where no good control variates are known. We illustrate our effectiveness in two production-grade exotic option-pricing problems, namely swaption with the Heath-Jarrow-Morton model and variance swap with a stochastic local volatility model.

4 - On Efficient Solutions of Structured Markov Processes in Quantum Computational Environments

Mark Squillante, IBM Research, Mathematical Sciences, Yorktown Heights, NY, United States

We study the fundamental problem of efficiently computing the stationary distribution of general classes of structured Markov processes. In strong contrast with previous work, we consider this problem within the context of quantum computational environments from a mathematical perspective and devise the first quantum algorithms for computing the stationary distribution of structured Markov processes. We derive a mathematical analysis of the computational properties of our quantum algorithms together with related theoretical results, establishing that our quantum algorithms provide the potential for significant computational improvements over that of the best-known classical algorithms in various settings of both theoretical and practical importance. Although motivated by structured Markov processes, our quantum algorithms have the potential for being exploited to address a much larger class of numerical computation problems.

MB44

Summit - 436

Simulation (tentative title)

Invited Session

Simulation Society

Chair: Wei Xie, Northeastern University, Boston, MA, United States

1 - Algorithms for Efficient Design of Simultaneous Controlled Experiments in the Presence of Subject Covariates

William Fisher, Clemson University, Clemson, SC, United States, Qiong Zhang

Algorithms for Efficient Design of Simultaneous Controlled Experiments in the Presence of Subject Covariates William Fisher and Qiong Zhang School of Mathematical and Statistical Sciences, Clemson University Abstract: Controlled experiments are a form of experiment that allows one to analyze a proposed treatment by comparing responses between a treatment group and control group. Researchers or organizations may run multiple, separate controlled experiments simultaneously due to time, budget, or other considerations. In this setting, subjects often participate in more than one experiment, and thus responses across experiments coming from the same subject are not independent. In addition, outside of the treatment, certain subject covariates may be explanatory factors in the subject's response to each experiment. In this talk, we consider the optimal experimental design problem of allocating subjects to treatment or control when subjects must participate in multiple controlled experiments simultaneously and subject covariate information is available. The goal of the allocation is to provide precise estimates of treatment effects for each experiment. Deriving the precision matrix of the treatment effects in this setting and using D-optimality as our allocation criterion, we propose algorithms which provide solutions to the D-optimality problem. These algorithms decompose the D-optimality problem into a sequence of subproblems, where each subproblem can be solved either through integer programming or a semi-definite programming based randomized algorithm that originates from the MAXCUT problem. We showcase the performance of our allocation algorithms and demonstrate other properties of the algorithms through a suite of simulation studies.

2 - Variance Reduction Based Experience Replay for Policy Optimization

Wei Xie, Northeastern University, Boston, MA, United States, Hua Zheng, Mingbin Feng

For reinforcement learning on complex stochastic systems, it is desirable to effectively leverage the information from historical samples collected in previous iterations to accelerate policy optimization. Classical experience replay treats all observations uniformly, neglecting their relative importance. To address this limitation, we introduce a novel Variance Reduction Experience Replay (VRER) framework, enabling the selective reuse of relevant samples to improve policy gradient estimation. VRER, as an adaptable method that can seamlessly integrate with different policy optimization algorithms, forms the foundation of our sample efficient off-policy learning algorithm known as Policy Gradient with VRER (PG-VRER). Furthermore, the lack of a rigorous understanding of the experience replay approach in the literature motivates us to introduce a novel theoretical framework that accounts for sample dependencies induced by Markovian noise and behavior policy interdependencies. This framework is then employed to analyze the finite-time convergence of the proposed PG-VRER algorithm, revealing a crucial bias-variance trade-off in policy gradient estimation: the reuse of older experience tends to introduce a larger bias while simultaneously reducing gradient estimation variance. Extensive experiments have shown that VRER offers a notable and consistent acceleration in learning optimal policies and enhances the performance of state-of-the-art (SOTA) policy optimization approaches.

3 - Efficient Nested Simulation for Risk Measurement: a Decoupling Approach

Nifei Lin, Fudan University, Shanghai, China, People's Republic of, Yingda Song, Jeff Hong

Estimating risk measures in real-time is essential for effective risk management but poses significant challenges, particularly when managing complex portfolios influenced by numerous risk factors. Traditional methods, relying on simplified models with analytical solutions, compromise accuracy for speed. Recent advances in nested simulations have enabled more realistic modeling, yet these methods are not sufficiently efficient for real-time application. We introduce a novel framework that decouples nested simulation into an offline learning phase for accurate loss surface fitting and an online application phase for rapid risk factor simulation. This approach significantly enhances computational efficiency and offers the flexibility to adapt various smoothing techniques for surface fitting. We demonstrate the framework's effectiveness by applying it to estimation of the systemic risk measure, CoVaR, a task that poses significant challenges under conventional nested simulation setups. Our results show that our method achieves superior convergence rates, presenting a significant advancement in real-time risk management.

4 - Langevin Diffusion-Based Linear Noise Approximation for Bioprocessing Mechanistic Model Bayesian Inference

Wandi Xu, Northeastern University, Boston, MA, United States, Wei Xie

Driven by the critical challenges in the bioprocess mechanistic and hybrid model inference, we propose a novel Bayesian inference approach called Langevin diffusion-base linear noise approximation (LD-LNA) to efficiently generate posterior samples. It is inspired by a combination of two mechanisms, i.e., Langevin diffusion from Metropolis-adjusted Langevin algorithm (MALA) and linear noise approximation (LNA). The Langevin diffusion leverages the gradient information from the target posterior to improve the mixing rate of the classic random walk-based MCMC sampling approach, while the LNA helps to bypass the difficult task of step size selection in the original MALA. We show that the convergence of the proposed LD-LNA to the desired posterior distribution under certain conditions. Further, we provide a quantitative bound to assess its finite-sample performance. Numerical experiments inspired by the real-world problems in the biomanufacturing processes demonstrate the effectiveness of the proposed algorithm, showing its potential to guide the most “informative” data collection to reduce model uncertainty and support optimal robust decision making.

5 - Development of subsampling methods for Multi-Fidelity optimization with constraints

Shun Tanaka, Osaka University, Suita, Japan, Hiroshi Morita

Multi-fidelity optimization is a method that integrates computational models of various complexities. It is suitable for situations where high-accuracy computations are costly, but time and costs can be reduced with low-accuracy models. For example, during the optimization of hyperparameters in deep learning, strategies for pruning and resource allocation can be considered by utilizing intermediate epoch evaluation values. By deciding the proportion of the total to be sampled and constructing a low-fidelity model, it is possible to accelerate the process over time. In this case, there is a trade-off between time and the optimal solution, so care must be taken. We have devised a method to reduce optimization time by generating different fidelity models through subsampling and incorporating the relationship between models into sequential model-based optimization.

MB45

Summit - 437

Innovative Interventions for Advancing Public Health

Invited Session

Health Applications Society

Chair: Amir Karimi, The University of Texas at San Antonio, San Antonio, TX, United States

1 - When Should Fractional-Dose Vaccines be Used?

Naireet Ghosh, London Business School, London, United Kingdom, Francis de Véricourt, Jérémie Gallien

Vaccination campaigns often face challenges such as limited stockpiles, delivery delays, and constrained administration capacity. In such contexts, fractional-dose vaccines have been described in the medical literature as a possible strategy because their efficacy reduction is typically less substantial than dose reduction, allowing for greater population coverage. We seek to determine the optimal use and potential benefits of a fractionated vaccine dose with lower and more uncertain efficacy, given the specific operational constraints faced in a vaccination campaign. We employ a Susceptible-Infected-Recovered (SIR) epidemic model integrating vaccination with full- and fractional-doses over time. We embed it within a deterministic optimal control model aimed at identifying vaccination policies that minimize total infections during an epidemic, taking into account restrictions on vaccine stockpile, delivery rate, and administration. Using a statistical approach from clinical literature for estimating the uncertainty around fractional-dose efficacy, we conduct two case-studies reflecting real-world scenarios. Our theoretical analysis provides an intuitive characterization of the optimal vaccination policy which, depending on the epidemic and operational parameters, may utilize a combination of full- and fractional-dose vaccines, either simultaneously or sequentially. We also examine simpler policies that employ a single vaccine dosage throughout the epidemic. Our findings suggest that, while single-dosage policies can be almost as effective as the optimal policy in averting infections, they are not as robust to the uncertainty affecting fractional-dose efficacies. We conclude that fractional-dose vaccines, used either alone or in conjunction with full-dose vaccines, present an opportunity to significantly reduce infections during an epidemic in resource-constrained settings.

2 - Strengthen Blood Transfusion System in Africa: Motivating Donations with Utility-Based Blood Drive Planning and Strategic Advertising

Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States, Bo Zeng, Jayant Rajgopal, Bopaya Bidanda, Pratap Kumar, Juan Puyana

Blood transfusion is a critical medical intervention that saves millions of lives annually. However, in many low- and middle-income countries, the blood supply often faces significant shortages with many deaths that actually are preventable. To address this critical issue, our presentation introduces a novel two-stage, multi-period optimization model designed to enhance the blood supply system through utility-based blood drive planning and strategic advertising. The model focuses on increasing the willingness to donate blood by incorporating various uncertainties associated with human behavior. It uniquely integrates decision-dependent uncertainty and ambiguity sets, allowing decision-makers to consider how their strategies may influence donor’s behavior over time. We will detail the model’s features, demonstrate its practical applications, and discuss its potential to improve blood availability and save lives in resource-limited settings.

3 - Modeling People's Decisions on Npi Compliance and Vaccination Willingness

Geonsik Yu, Purdue University, West Lafayette, IN, United States, Michael Garee, Mario Ventresca, Yuehwern Yih

Protecting public health from infectious diseases often requires a collective effort from community members. Individual decisions regarding vaccination, hand-washing, social distancing, and mask-wearing can collectively have an external effect or synergy in disease mitigation. The challenge is that the information individuals use for their decision-making may be limited due to their choice of information sources. Additionally, their reasonable decisions, based on situations such as resource availability, can inadvertently lead to outcomes unfavorable to the entire system. Recognizing the crucial role of each individual's choices is paramount to ensuring the overall effectiveness of public health measures.

To understand how people make decisions in response to a spreading disease, we developed decision-making scenarios from an individual’s perspective based on insurance theory and sequential decision trees. We then designed the corresponding agent-based simulation framework to model the consequences of individual decisions and examine the effects of various opinion interventions for disease mitigation. Our focus is on people's decisions in the context of a pandemic, where they can choose either to wear a mask or not and either to get vaccinated or not, aiming to minimize potential losses due to infectious diseases. We calibrated the agent-based simulation model to the empirical data of the COVID-19 pandemic to generate realistic testbeds for our study. Additional datasets were also incorporated to capture some characteristics of the actual population. Our tentative results highlight the importance of understanding the characteristics of the audience population when choosing channels for opinion intervention for disease mitigation.

4 - The Role of Market Interventions to Increase Access of Health Commodity in Lmic: A Case of The Ipaqt

Parshuram Hotkar, Indian School of Business, Hyderabad, India, Sarang Deo

Low and medium income countries (LMICs) face challenges to provide access to health commodities. This paper explores the use of market interventions that coordinate pricing in a supply chain to improve access to diagnostic tests for tuberculosis (TB) in the LMICs. Specifically, we analyze the role of the Initiative for Promoting Affordable and Quality TB Tests (IPAQT) in India, which aimed to lower the cost of the Xpert MTB/Rif assay through negotiations with manufacturers and coordination with private labs. Using a game-theoretic model, we assess the feasibility and sustainability of market-transforming interventions and their impact on test consumption. We identify the conditions under which such interventions successfully coordinate the transformation from a high-price, low-volume market to a low-price, high-volume market. Additionally, we demonstrate that internal price control mechanisms like IPAQT lead to higher consumption compared to external mechanisms such as price regulations, particularly in the presence of heterogeneity in private labs. This research contributes to our understanding of the economics of market-transforming interventions, their generalizability, and provides guidelines for policy makers and donors for intervention designs aimed at enhancing access to health commodities in the LMICs.

MB46

Summit - 438

HAS Distinguished Lecturers Session

Award Session

Health Applications Society

Chair: Lauren Steimle, Georgia Tech ISyE, Atlanta, GA, United States

Co-Chair: Hrayer Aprahamian, Texas A&M University, College Station, TX, United States

1 - Optimization of Active Surveillance for Prostate Cancer Under Model Ambiguity

Brian Denton, University of Michigan, Ann Arbor, MI, United States, Weiyu Li

Variation across longitudinal medical studies can lead to multiple stochastic models of disease progression. To address this ambiguity, we propose a new multi-model partially observable Markov decision process (MPOMDP) framework. Here, model ambiguity is defined as the case where there are multiple credible optimization models with the same structure but different model parameters. The proposed MPOMDP model aims to learn the distribution of the true model from system outputs over time and to find the single optimal policy that maximizes the expected sum of all future rewards in all possible models. We discuss properties of the proposed MPOMDP model which motivate solution methods for MPOMDPs. We develop an exact solution method and two approximation methods that are shown to converge asymptotically and compare their performance in computational experiments. Finally, we use a case study of active surveillance for prostate cancer to demonstrate how the MPOMDP model can be applied to a real-world problem to improve medical decision-making by creating policies robust to model ambiguity.

2 - From Kidney Allocation to Administrative Costs: My Health Care OR Journey

Stefanos Zenios, Stanford University, Stanford, CA, United States

In this talk I will share the lessons I have learned over 30 years of Health Care OR. I will focus on 4 specific projects: a) Managing the equity-efficiency trade-off in health care; b) Using incentives to improve healthcare delivery; c) Understanding product development in medical devices; d) Reducing administrative costs. There will be some equations and some philosophical reflections.

3 - AI for Social Good - Leveraging Digital Therapeutics and Advanced Analytics for Health Literacy

Rema Padman, Carnegie Mellon University, Pittsburgh, PA, United States

Health literacy is a widely recognized challenge worldwide, with many adults lacking the requisite skills to engage successfully in the management of their health and healthcare. Affecting both individual and societal health outcomes, limited health literacy particularly exacerbates the increasing physical and psychological burden for patients with multiple health conditions as well as pediatric, elderly and disadvantaged populations. Recent developments in digital therapeutic solutions offer an opportunity to apply systems thinking combined with analytic methods to synthesize the myriad components of a multi-pronged approach to improving societal health literacy at scale. This talk will highlight some of these developments with a focus on digital platforms and algorithmic artifacts in the healthcare delivery setting, recognizing the challenges of misinformation, disinformation, inclusivity and representativeness in identifying and disseminating authoritative and accurate content for educating and empowering patients and the public.

MB47

Summit - 439

Multimodal AI for Medicine and Hospital Operations

Invited Session

Health Applications Society

Chair: Dimitris Bertsimas, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Kimberly Villalobos Carballo, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Lisa Everest, Operations Research Center, Massachusetts Institute of Technology, Cambridge, United States

1 - Prediction of Left Ventricular Ejection Fraction from a 12-Lead Electrocardiogram

Catherine Ning, MIT, Cambridge, MA, United States, Dimitris Bertsimas, Yu Ma

Within the Holistic AI in Medicine (HAIM) framework, we propose a promising opportunity to develop a more efficient and cost-effective approach to estimate left ventricular ejection fraction (LVEF) using 12-lead electrocardiogram (EKG) data. LVEF is one of the essential parameters for assessing cardiac function as it represents the percentage of blood ejected from the heart's left ventricle during each contraction (the normal range is 55-65%). A reduced LVEF may serve as an crucial diagnostic and prognostic marker for heart failure severity and risk of sudden death. Traditionally, EF estimation relies on time-consuming and expensive imaging techniques, such as echocardiography or MRI. In our project, we aim to extract accurate EF predictions from routine, readily available EKGs only, which could significantly enhance early detection, risk stratification, and treatment monitoring for patients both with and without known history of heart failure.

In close collaboration with the cardiologist team at Hartford Healthcare (HHC), we carry out a retrospective study on outpatients from the HHC data systems who have both a 12-lead EKG and an echocardiogram record, the latter serving as the ground-truth LVEF measurement. Our preliminary 4-class classification results based on 12 leads attain a per-class AUC averaging to 0.883 using XGBoost models. We are excited to extend this work to a multi-modal, multi-task unified prediction framework targeting other heart-related problems such as hypertrophic cardiomyopathy, aortic stenosis, and atrial fibrillation from EKGs alone.

2 - From Data to Action: HAIM Modeling for Medication-Related Readmissions

Yubing Cui, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, Lisa Everest

Prior to discharging an inpatient, it is optimal for pharmacists to review each patient's medications to mitigate the risk of medication-related complications. However, limited pharmacist resources often render this impractical. In collaboration with Hartford HealthCare and leveraging the HAIM framework, we are developing a multimodal machine learning model designed to identify patients at high risk for medication-related adverse events or readmissions. This approach allows for the strategic allocation of pharmacist resources to those patients who are most in need.

3 - Automating care reviews for denial management with multimodal AI

Karl Zhu, Massachusetts Institute of Technology, Cambridge, MA, United States, Yubing Cui, Dimitris Bertsimas

We are partnering with Hartford HealthCare, Connecticut's largest healthcare system, to develop an automated care review system with multimodal AI. This initiative aims to mitigate errors in level-of-care decisions and reduce insurance denials, which impose significant financial challenges within the healthcare sector. Traditional care review methods, reliant on manual data extraction from electronic health records, are inefficient and prone to inaccuracies, frequently resulting in unjustified denials.

Our multimodal approach combines structured tabular data — including patient demographics, vitals, lab results, and medications — with embeddings derived from unstructured textual data, such as provider notes. Preliminary tests on classifying patient status for those admitted through the emergency department demonstrate that our multimodal approach outperforms unimodal ensemble methods in terms of AUC.

4 - Multimodal Approach to Predicting Psychosis Relapse

Emily Hahn, Massachusetts Institute of Technology, Cambridge, MA, United States, Jiayi Gu

Predicting prognostic outcomes (length of stay, readmission,etc.) represents a crucial challenge in psychiatric healthcare. For example, relapse in psychosis is often preceded by detectable symptom changes nearly two months in advance. However, current screening tools have limited predictive capability. Aiming to enhance patient outcomes and reduce economic burdens on health systems, this study utilizes the Holistic Artificial Intelligence in Medicine (HAIM) framework to develop advanced machine learning algorithms using multimodal EMR data for predicting prognostic outcomes such as re-admission . This retrospective study includes 6800 adult patients admitted to the Behavioral Health Network within Hartford Healthcare from 2018 to 2023, Our models integrate advanced machine learning algorithms to process structured and unstructured EMR data, including clinical notes through Natural Language Processing (NLP) and brain imaging when available. By creating predictive models using the HAIM framework, the study aims to demonstrate the potential of such tools in improving prognoses and long-term outcomes for patients, ultimately supporting more proactive and personalized healthcare strategies.

MB48

Summit - 440

Managing Product Returns

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Huseyn Abdulla, University of Tennessee, Knoxville, TN, United States

Co-Chair: Nail Tahirov, University of Zurich, Zurich, Switzerland

1 - Salesperson Incentives and Consumer Returns

Seulchan Lee, Michigan Technological University, Houghton, MI, United States, Huseyn Abdulla

As a widely-adopted practice, retailers deduct from the paychecks of salespeople entire sales commission earned on sales of products that are eventually returned to stores, for any reasons. This practice, coupled with lenient no-questions-asked, full-refund policies, results in significant paycheck uncertainty for salespeople and threatens their economic welfare. We examine the monetary refund policies and commission chargeback decisions of a retailer in response to consumer product returns. Our results challenge the widely adopted practice of "full refund for customers, full commission chargeback from salespeople."

2 - Pricing and Returns in the Era of Big Tech: Implications of Information Asymmetry Reversal

Kiarash M. Hassani, Queen's University, Canada, Kingston, ON, Canada, Murray Lei, Anton Ovchinnikov

We present a model to optimize the sales and return policy of a monopolistic seller under three information scenarios and examine the resultant profit and consumer surplus. The consumer's ex-ante information is her type and the firm's ex-ante information about a consumer may be the distribution of consumer types (uninformed firm, standard information asymmetry), the consumer type (equally-informed firm), or the actual validation (better/fully-informed firm, reversal of information asymmetry). In the base model we assume consumers incur no hassle costs for returns, and relax this in an extension.

The results of the base model show that when returns are not allowed, the firm collects twice as much profit when it has equal information compared with the standard information asymmetry but being fully informed brings no additional value. When returns are allowed, full information is always better than equal information, and equal information is always better than no information for the firm. Likewise, when the firm is fully informed, it is always profitable to allow returns. The value of information is decreasing in salvage value, and when the salvage value equals the production cost, additional information does not improve the profit. When returns are not free for consumers, we end up with various pricing policies based on the magnitude of return hassle costs and other variables. These policies and their effects on the firm and consumers differ depending on the firm's information level.

3 - Does Less Result in More? the Role of Information and Rewards in Take-Back Programs for Clothing

Anna Saez de Tejada Cuenca, IESE Business School, Barcelona, Spain, Erin Mckie, Vishal Agrawal

Fashion retailers are increasingly implementing take-back programs to reduce textile waste and prevent clothes from being landfilled. Through a series of experiments involving over 3,500 participants, we examine the effect of different types of information and reward levels on consumers' propensity to return their used clothes to take-back programs.

4 - The Impact of *Buy Now, Pay Later* on Customer Sales and Returns in Online Retail

Christoph Baldauf, Stockholm School of Economics, Stockholm, Sweden

Product returns remain a major issue for online retailers. The emergence of new, flexible, payment methods that allow customers to buy products without having to pay at the point of purchase may incentivize them to return products at an even higher rate. Such forms of payment are commonly referred to as “Buy Now, Pay Later” (BNPL). In this paper, we examine the role of customers’ payment method choice in retail transactions by focusing on the impact of BNPL on sales and returns in online retail. We analyze a unique dataset from a global fashion retailer containing over 30 million transactions from two European markets. Our findings suggest that, compared with other payment methods, BNPL increases sales, but also leads to a substantial increase in returns. We dive into factors that may moderate the negative effect of BNPL on returns and conduct back-of-the-envelope analyses to quantify its economic impact.

MB49

Summit - 441

Empirical Platform Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yifan Feng, National University of Singapore, Singapore, Singapore

1 - Impact of Estimated TIME of Arrival on the Supply and Demand for Food Delivery Platforms

Qianyi Tao, Tsinghua University, Beijing, China, People's Republic of, Tirui Cao, Lei Zhao, Hai Wang

In food delivery services, the *estimated time of arrival* (ETA) is a crucial decision that may significantly impact the demand and supply. Serving as a service commitment to customers and a delivery requirement for couriers, ETA interconnects customers (the demand side) and couriers (the supply side). Therefore, the design of the ETA strategy may simultaneously influence the satisfaction and retention of both customers and couriers. This research aims to study the causal effect of ETA on customer satisfaction, customer retention, and courier retention. To address potential endogeneity issues, we use the Heckman two-stage method to control the potential sample self-selection bias in the customer satisfaction model. Additionally, we meticulously identify appropriate explained variables, select explanatory variables, and introduce proxy variables into our econometric models. Based on the real data from a major food delivery platform in China, we discover that, as ETA increases from the current status quo in Beijing, the customer dissatisfaction rate will first decrease to a minimum point and then increase, the customer retention will decrease, and the courier retention will increase. Robustness checks across various cities and time slots validate our empirical results. Furthermore, a large-scale randomized field experiment conducted on the demand side of our collaborator's platform also validates the credibility of our models. Interestingly, the experimental results show that an increase in ETA corresponds to a subsequent increase in the *actual time of arrival* (ATA). Our study sheds valuable managerial insights into designing ETA strategies for food delivery platforms.

2 - Optimal Stockist Selection and Contract Design: Evidence from a Supply Chain in India

Yu Long, the Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Jussi Keppo, Wei Jiang, Omkar D. Palsule-Desai

We study incentives and contract design when a manufacturer can select, contract, and learn about its stockists. We model this relationship using a two-period principal-agent framework in which the manufacturer is the principal and the stockists are agents. Using a multi-armed bandit model, we examine how competition among stockists for future contracts and their career concerns on the level of future compensation influence the optimal contract. We show that competition incentivizes stockists to exert more effort, whereas career concerns have opposite incentives, particularly when the manufacturer has dominant bargaining power. By calibrating our model to unique contract and sales data from an Indian potato chips manufacturer, our counterfactual analysis shows that the competition raises (career concerns decrease) the stockists' effort by more than 192% (77%) and the manufacturer's value by over 6% (about 3%).

3 - Unlocking Nanostores' Operational Potential via a Buy Now Pay Later Program

Guan-Yuan Wang, Tilburg University, Tilburg, Netherlands, Jan Fransoo, Han Oh

We study how Buy Now Pay Later (BNPL) trade credit can foster business performance for Base-of-Pyramid retail shopkeepers, leveraging granular data from East Africa. By employing difference-in-difference estimators and quantile regression models, we show that BNPL credit can improve BOP retailers' business growth. By estimating different estimands, we suggest inclusive BNPL policies would greatly benefit small-scale retail stores.

4 - Understanding Labor Supply in Gig Economy: Evidence from a Logistics Platform

Sixing Hu, National University of Singapore, Singapore, Singapore, Yifan Feng, Jussi Keppo

We provide a detailed empirical examination of labor supply behavior in an e-commerce logistics platform, contributing to the nuanced interplay between daily compensation and labor supply. Using a unique dataset, which contains information about predetermined compensation for ``scheduled'' jobs and compensation offered to non-participating agents, we avoid many common sources for endogeneity. We employ a collection of econometrics tools to empirically characterize the daily extensive (i.e., whether the agent participates) and intensive (i.e., how long the agent works) margins, respectively. We also investigate the ``intraday'' dynamic margins, meaning how compensation changes of times of the day affect the labor supply in the remaining of the day.

Our findings regarding the intensive margins confirm the qualitative insights of Camerer et al. Specifically, we find that the intensive margin is negative. Our intraday analysis finds that the negative elasticity mainly comes from near the end of the day, supporting the interpretation that agents set ``income targets.'' However, the intensive margin is only part of the story since the agents in the gig economy do not typically work full-time. We find that the extensive margin is positive. That is, higher compensation leads to more agents to participate on a given day. The positive elasticity mainly comes from the beginning of the day, suggesting that agents first decide on whether to work and then how long to work. After combining intensive and extensive margins, we find that the total labor supply correlates positively with compensation, thus reverting the qualitative takeaway of Camerer et al..

5 - Bargaining in Parallel: the Impact of Multiple Concurrent Bargaining on Final Selling Prices

Wen Zhang, Baylor University, Waco, TX, United States, Guihua Wang

Many studies suggest that competition among buyers typically benefits sellers. However, this may not always hold true in online marketplace bargaining when multiple buyers simultaneously compete for the same item. In this paper, we examine the impact of multiple concurrent negotiations on the selling prices of items. Utilizing a novel dataset, we find that an increased number of concurrent bargaining instances can actually decrease an item's selling price. Furthermore, we employ tree-based machine learning algorithms to explore the heterogeneous effects of competition among buyers on bargaining outcomes. This analysis provides practical insights for online marketplace sellers.

MB50

Summit - 442

Empirical Research in Supply-Chain Networks

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Kashish Arora, Indian School of Business, Hyderabad, India

1 - Supply Chain Cash Flow Chess: Seasonality in Trade Credit Late Payment

Zhi Li, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Jing Wu, Ling Cen

Using a sample of US public firms and a unique dataset of trade credit payments, we document several key findings regarding this late payment seasonality. We observe spikes in trade credit late payments at the fiscal year-end. We also find such late payment in trade credit is associated with an increase in customers' cost of goods sold, suppliers' sales, and both of their cash ratios at fiscal year-end. However, the focal firms do not increase their accounts payable, suggesting that they are only delaying payments for existing trade credit instead of receiving more credit. These results suggest delayed payments are related to changes in working capital around fiscal year-end. We find firms with higher short-term debt are more likely to delay payments, suggesting trade credit functions as a substitute for bank financing. Late payment seasonality is also more pronounced for firms facing tighter financial constraints. However, we find no evidence that late payment could hurt either suppliers' or customers' profits. Our findings shed new light on why and how trade credit late payment seasonality arises and its implications for firms' financial flexibility and liquidity management.

2 - Vertical Integration and Market Power in Supply Networks

Kashish Arora, Indian School of Business, Hyderabad, India, Amandeep Singh, Mamta Sahare

Vertical integration, a strategic approach where firms control multiple stages of production and distribution, offers potential operational efficiencies. While the operations management literature emphasizes its benefits for supply chain coordination, and reduced transaction costs, certain antitrust studies concern how vertical mergers may enable dominant firms to limit competition by creating entry barriers and causing market foreclosure. There is some evidence of the potential supply chain benefits of vertical integration across a few industries. However, empirical evidence on its anti-competitive effects remains scarce. This paper contributes to closing this gap by examining vertical integration's impact on firms' market power, as measured by their price-marginal cost markups. Our research investigates pre- and post-integration markups to determine whether firms gain anti-competitive market power after integration. Since firms' marginal costs are not usually observed in data, we estimate markups using production function estimations, following recent advances in empirical industrial organization. We compile a novel dataset on vertical integration cases using the FactSet Revere and the SDC platinum databases. Specifically, we identify 213 vertical mergers between firms with existing buyer-supplier relationships from 2003 to 2022. We employ a staggered difference-in-difference approach in conjunction with instrumental variables that cause an exogenous variation in firms' decisions to integrate vertically. Our analysis reveals that vertical integration increases acquiring firms' markups by 13\% and their rivals' markups by 12\%. Interestingly, we find that the increase in firms' market power post-integration does not come from improved operational efficiencies but from higher than competitive prices.

3 - Examining the Evidence of Value Smile Curve

Nikolay Osadchiy, Emory University, Atlanta, GA, United States

In this article, we examine the existence of Value Chain Smile, a concept first introduced by Stan Shih, then CEO of Acer group, at the firm level for all U.S. companies and over time. We study whether the most upstream or downstream firms capture more value compared to the intermediate firms. We further investigate the value addition at the industry level and over time. In the last, we study whether the firms that move upstream or downstream can extract more value following the evidence of the smile curve.

4 - A framework for synthetic graph generation with user-specified metrics

Nolan Feeny, University of Michigan, Ann Arbor, MI, United States, Seth Guikema

Supply chain managers have historically prioritized manufacturing efficiency at the cost of greater scale supply chain network (SCN) resiliency. Although researchers have recently put more emphasis into understanding SCN network resiliency, lack of substantial data remains a common obstacle due to competition and security reasons. This presentation proposes a framework to generate synthetic networks exhibiting network metric targets within an error range that users can define and adjust. We combine existing graph generation algorithms with a genetic optimization algorithm focused on specific network-defining metrics to score each graph. Metrics such as average nodal degree, average shortest path length, and clustering coefficient are indicative of network resilience and are thus included in our scoring function. The resulting output of the genetic algorithm is a final network that meets the user's desired specifications. The framework demonstrates an ability to successfully generate networks with metrics resemblant of real-world networks with user-defined metric targets and precision. This is an alternative method to provide sufficient and realistic synthetic data for network resilience analysis and machine learning applications, especially in industries where data is difficult to obtain.

MB51

Summit - 443

Research at iFORM with Social Considerations

Invited Session

MSOM: iForm

Chair: Xin Geng, University of Miami, Coral Gables, FL, United States

Co-Chair: Xiaomeng GUO, Hong Kong Polytechnic University, Kowloon, Hong Kong

1 - How Sustainable are Sustainability-Linked Loans?

Jonathan Hsu, The University of Hong Kong, Hong Kong, Hong Kong, Wenqiang Xiao, Guoming Lai

Sustainability-linked loans (SLLs) have grown to rival green loans as one of the principal instruments of green financing. SLLs eschew the use-of-funds model used by green loans in favor of a performance-based model wherein the repayment terms are linked to the sustainability performance of the firm. However, it is unclear whether and how SLLs may be more effective than green loans at incentivizing ESG improvements by the borrower. We develop a model of corporate financing wherein a borrower with a green project can improve pecuniary and non-pecuniary outcomes as well as reduce their default risk by investing effort into improving their ESG performance. We show that a green loan results in ESG underinvestment due to both the inability to align lender and borrower preferences over non-pecuniary outcomes and the endogeneity of borrower default. An SLL eliminates these agency frictions by tying the loan repayment terms to the borrower's *ex-post* ESG performance, thus restoring the first-best allocation of sustainability effort. This result holds even when the ESG performance of the borrower can only be measured imprecisely by the lender. However, noisier measurements "mask" the true ESG performance of the borrower, forcing the lender to issue more performance-sensitive SLLs in equilibrium. Finally, through an analysis of an information asymmetric setting, we show that an SLL can function as an inexpensive screening device. In particular, by committing to more ambitious sustainability targets, borrowers can credibly reveal private information about the effectiveness of their ESG performance at mitigating default risk.

2 - Price Competition in the Presence of Social Comparison and Demand Uncertainty

Yun Zhou, McMaster University, Hamilton, ON, Canada, Zhoupeng Zhang, Ming Hu, Tony Cui

Does social comparison intensify the competition among sellers in volatile markets? In this paper, we study price competition in markets with volatile sizes and, more importantly, with sellers that socially compare their profits with each other ex post for any realization of market sizes. While the classic social comparison theory, as well as conventional wisdom, has long suggested that the comparison behavior such as behind aversion (upward comparison) and ahead seeking (downward comparison) will intensify competition, we reveal how social comparisons can totally have opposite-directional impacts once interacting with the market variability. In particular, we show that a stronger behind aversion behavior will still intensify the price competition, yet somewhat surprisingly, the impact of ahead seeking is variability-dependent: there is a threshold on the market size variability such that the competition between sellers will, in fact, be softened if markets are more volatile than such a threshold, and vice versa. Interestingly, the aforementioned predictions go in the other direction when sellers are selling complementary products: we will have the impacts of behind aversion variability-dependent while the ahead seeking is always pro-competitive. Our insights are robust under a general demand structure and demand shock specifications, asymmetry in sellers and markets, and sellers' misperceptions in market variabilities. Our work sheds light on other interactions of strategic complements or substitutes in the presence of social comparisons, e.g., quantity competition and advertising competition.

3 - Impact of Direct Selling on Food Waste

Xinxiu Wang, University of Miami, Miami, FL, United States, Xin Geng, Nan Yang

The yield of agricultural produce is random and affected by many uncontrollable factors, which stimulates farmers' overproduction, ultimately leading to a large amount of food waste. Direct marketing is thought to be an effective strategy for reducing food waste by disposing of excess produce through direct sales channels. However, the benefits that direct selling brings to farmers may incentive them to produce more, thereby becoming a potential cause of increased waste. Motivated by such trade-offs, we establish a game-theoretic model to study the impact of direct selling on food waste.

4 - Product Bundling in the Presence of Vertical Differentiation

Amit Eynan, University of Richmond, Richmond, VA, United States, Xingxing Chen

We study a seller of two product types which can be sold as individual components or at a discount as a bundle in order to leverage customers' product preference variation as reflected through their willingness to pay. Because customers are also heterogenous in their affinity towards quality, we consider offering each product type in two quality versions, high and low. This results in multiple potential bundling offerings across the two product types. As each customer self-selects a utility maximizing option, we investigate which components or bundles the seller should offer based on the market characteristics and highlight the impact of combining these two discrimination techniques on product quality, price, and profit.

MB52

Summit - 444

Causal Inference for Dynamic Operations

Invited Session

MSOM: Service Operations

Chair: Stefan Wager, Stanford GSB, Stanford, CA, United States

1 - Switchback Experiments under Geometric Mixing

Yuchen Hu, Stanford University, Stanford, CA, United States, Stefan Wager

The switchback is an experimental design that measures treatment effects by repeatedly turning an intervention on and off. While the switchback is a robust way to overcome cross-unit spillovers, it is vulnerable to bias from temporal carryovers. In this paper, we consider properties of switchback experiments in Markovian systems that mix at a geometric rate. We find that, in this setting, standard switchback designs suffer considerably from carryover bias: Their estimation error decays slowly in terms of the experiment horizon, whereas in the absence of carryovers a faster rate would have been possible. We also show, however, that judicious use of burn-in periods considerably improves the situation, and enables errors that decay almost as fast as a parametric rate. Our formal results are mirrored in an empirical evaluation.

2 - Posterior Sampling by Autoregressive Generation

Daniel Russo, Columbia University, New York, NY, United States

We provide a bridge between generative sequence modeling and sequential decision making problems that require quantifying uncertainty and judiciously exploring to resolve it. We demonstrate ideas through a meta bandit modeling the recurring problem of learning about newly released content on a recommender system. Our solution first pre-trains an autoregressive model to predict a sequence of repeated feedback/outcomes (e.g., responses to news articles shown to different users over time). At decision-time, we autoregressively sample an imagined sequence of outcomes for each action, and choose the action with the largest average imagined reward. Far from a heuristic, our approach is an implementation of Thompson sampling (with a learned prior), a prominent active exploration algorithm. We prove our pretraining loss directly controls online decision-making performance. Then, we demonstrates our insights empirically on a news recommendation task; this demonstration includes end-to-end fine-tuning of a language model and synthesizing it with data on user interactions.

3 - Balancing Risk and Reward: A Batched-Bandit Strategy for Automated Phased Release

Iavor Bojinov, Harvard Business School, Somerville, MA, United States, Yufan Li, Jialiang Mao

Phased releases are a common strategy in the technology industry for gradually releasing new products or updates through a sequence of A/B tests in which the number of treated units gradually grows until full deployment or deprecation. Performing phased releases in a principled way requires selecting the proportion of units assigned to the new release in a way that balances the risk of an adverse effect with the need to iterate and learn from the experiment rapidly. In this paper, we formalize this problem and propose an algorithm that automatically determines the release percentage at each stage in the schedule, balancing the need to control risk while maximizing ramp-up speed. Our framework models the challenge as a constrained batched bandit problem that ensures that our pre-specified experimental budget is not depleted with high probability. Our proposed algorithm leverages an adaptive Bayesian approach in which the maximal number of units assigned to the treatment is determined by the posterior distribution, ensuring that the probability of depleting the remaining budget is low. Notably, our approach analytically solves the ramp sizes by inverting probability bounds, eliminating the need for challenging rare-event Monte Carlo simulation. It only requires computing means and variances of outcome subsets, making it highly efficient and parallelizable.

MB53

Summit - 445

Modeling and Analytics in Humanitarian Logistics

Invited Session

MSOM: Supply Chain

Chair: Naren Agrawal, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, 95053, United States

1 - Enhancing Effectiveness and Responsiveness of Humanitarian Food Aid Delivery &Mdash;&Mdash; Data-Driven Optimization Framework

(Lluvia)Weijia Jing, Northeastern University, Boston, MA, United States, Ozlem Ergun

Each year, the Bureau for Humanitarian Assistance (BHA) under the U.S. Agency for International Development (USAID) plays a vital role in the global distribution of food aid. However, meeting the increasing demand with finite resources remains a persistent challenge(ref). In this paper, we examine the USAID global food aid supply chain, focusing on its multi-echelon, multi-commodity, and multi-modal nature. We adopt a multi-stage stochastic programming model to optimize warehouse locations and safety stock levels. This model integrates pre-disaster pre-positioning with post-disaster inventory replenishment, effectively catering to both deterministic (ongoing) and stochastic (sudden-onset) demands. Leveraging advance deep learning models, our approach enhances demand pattern analysis and realistic scenario planning. We aim to optimize BHA/FFP's food aid procurement, prepositioning, and distribution strategies, thereby enhancing the resiliency and responsiveness of USAID's global food aid operations.

2 - A Location-Allocation Model for Meal Kit Distribution in Post-Disaster Environments

Diana Ramirez-Rios, University at Buffalo, Buffalo, NY, United States, Angelo Soto-Vergel, Trilce Encarnacion, Johanna Amaya

This presentation focuses on the food bank network design in post-disaster environments that minimizes the social costs of human suffering. The social costs include the costs of the delivery operation and the external costs on the population's suffering caused by the delays in the relief in the form of deprivation costs. We propose a location-allocation model that selects the optimal location and allocation of PODs (i.e., food pantries) for the distribution of food after an emergency occurs, as well as the relief distribution strategy. Given a set of food bank locations, a set of PODs considered for the distribution of food supplied by the food bank, and a set of population regions (e.g., ZIP codes), the model searches for the optimal PODs that can be activated to serve the people in need, their allocation to the population centers, and the amount and frequency of truck deliveries to the POD during the critical period of the emergency. We study the case of the Houston Food Bank and its response after Hurricane Harvey with the actual network of PODs and its capacities. The deprivation cost function used was empirically estimated for meal kit distribution and is a polynomial function with respect to time that people have to drive and wait for the delivery.

3 - D-optimal orienteering for post-earthquake reconnaissance planning

Ilya Ryzhov, University of Maryland, College Park, MD, United States, Jiaqi Wang, Weijun Xie, Nikola Markovic, Gaby Ou

Immediately following a major earthquake, reconnaissance surveys seek to assess structural damage throughout the region with the help of a limited number of on-ground inspections. The goal is to collect informative and representative data that will guide subsequent relief efforts. We formulate a new type of vehicle routing problem, in which vehicles are tasked with data collection, and the objective function measures data quality using a nonlinear, nonseparable experimental design criterion. We create novel exact methods for this problem, and demonstrate their practical potential in a realistic case study using a state-of-the-art earthquake simulator.

4 - The Price of Funding Inflexibility in Humanitarian Operations

Ece Gulserliler, Tilburg University, Tilburg, Netherlands, Atalay Atasu, Dan Iancu, Luk Van Wassenhove

Humanitarian organizations (HO) respond to many emergencies with limited financial resources. For an effective and timely response, they need to prepare for potential emergencies, and allocate resources efficiently between programs. However, current funding practices often impede this. HOs receive a significant portion of funds in the aftermath of a crisis, preventing preparedness efforts. Moreover, these funds are often earmarked to a specific response, preventing organizations from preparing for and responding to other (usually smaller, under-the-radar) crises. This paper introduces a framework to quantify the price (i.e. value loss to beneficiaries) of such donor-imposed constraints in humanitarian operations. To that aim, we use a two-stage stochastic model where in the first stage, a humanitarian organization makes preparedness investments for multiple potential emergencies with unknown demand, subject to a budget and constraints on the use of the budget. In the second stage, the demands are realized and the HO utilizes the preparedness investments to respond. We quantify and compare the prices due to earmarking and preparedness constraints, and discuss how they depend on the demand and cost parameters. Our results have important implications for both humanitarian organizations and donors for understanding the consequences of funding inflexibility in various contexts.

5 - The Whiplash Effect: Congestion Dissipation in a Disrupted Circulatory System

Chaoyu Zhang, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu

In this paper, we build a fluid model in transient states to investigate how disruptions at one port can affect the disrupted port and its counterpart in a circulatory container shipping system. Port disruption leads to two types of congestion: the inbound backlogs, which occur when ships are unable to enter the port, and the outbound backlogs, which arise when goods are unable to be loaded onto ships. We characterize the recovery time of the system of two ports in closed form and the dynamics of inbound and outbound backlogs during the recovery process. We identify a whiplash effect in the outbound backlog process at both ports in which ports take turns being whipsawed by outbound backlogs that may go up and down. This counterintuitive phenomenon is due to the uneven spatial distribution of ships on the routes caused by the disruption. Furthermore, using the base two-port model as a building block, we obtain more general results in characterizing the recovery process of those bipartite port networks. We also extend the fluid model to a diffusion model to capture the second-order effect of the real-life system. Finally, we use our model prediction as a feature and apply machine learning techniques to predict the time that vessels spend in the Shanghai port after a typhoon disruption and show that using our model predictor, alongside all relevant vessel and voyage attributes as raw features, significantly reduces prediction errors by 10% to 30% on average compared to not using it.

MB54

Summit - 446

MSOM TIE Panel

Panel Session

MSOM:Technology, Innovation, and Entrepreneurship

Co-Chair: Sidika Candogan, National University of Singapore, Singapore, Singapore

Co-Chair: Antoine Feylessoufi, University College London, London, United Kingdom

1 - Moderator Panelist

Antoine Feylessoufi, University College London, London, United Kingdom

2 - Panelist

Stylianos Kavadias, University of Cambridge, Cambrige, United Kingdom

3 - Panelist

Christoph Loch, Cambridge University, Cambridge, United Kingdom

4 - Panelist

Matthew Marx, Cornell University, Ithaca, NY, United States

5 - Panelist

Jurgen Mihm, Insead, Fontainebleau, France

6 - Panelist

Kamalini Ramdas, London Business School, London, United Kingdom

7 - Moderator Panelist

Sidika Candogan, National University of Singapore, Singapore, Singapore

MB55

Summit - 447

Healthcare Operations

Invited Session

Public Sector OR

Chair: Hessam Bavafa, Wisconsin School of Business, Madison, WI, United States

1 - Emergency Department Wait TIME Forecasts: Outdoing Complex Machine Learning with Parametric Models

Donald Lee, Emory University, Atlanta, GA, United States, Ragip Gurlek, Francis de Vericourt

Patient time-to-treatment is a key metric for benchmarking the operational performances of emergency departments (ED). In the past decade, increasingly sophisticated machine learning methodologies have been developed to better forecast wait times for individual patients. In this work, we explore how using information specific to the ED setting leads to simpler models that outperform more complex machine learning approaches.

2 - Primary Care Panel Optimization

Dilara Sonmez, The University of Chicago Booth School of Business, Chicago, IL, United States, Daniel Adelman, George Weyer

We target primary care providers' success in value based care through altering primary care visit patterns. We formulate the problem of finding physicians' optimal patient panels and prescribing patients with a primary care dosage as a mixed-integer nonlinear optimization problem. We take into account the operational constraints and primary care visit pattern dynamics. We capture the effect of visit patterns on the financials and incorporate those dynamics into the optimization model. We propose a heuristic method to be used for the large instances of the problem.

3 - Failures of Health Equity: An Examination of Bias in Clinical Trials

Ling Li, Indiana University, Bloomington, IN, United States, Kurt Bretthauer, Xiaoshan Peng

Failures of health equity are closely intertwined with biases in healthcare treatment. We study how underrepresentation bias in clinical trials affects health outcomes and equity. We propose behavioral models to analyze physicians’ treatment decisions. In our model, as supported in the literature, physicians’ treatment decisions are influenced by two distinct cognitive systems: the intuitive and the rational. We model the intuitive system as a Markov decision process where decisions are influenced by the treatment outcome of the previous patient using a "win-stay/lose-shift" heuristic. This illustrates a propensity among physicians to modify their treatment choices following negative outcomes. By modeling the rational system as a multi-armed bandit problem, physicians select treatments based on their subjective probabilities of treatment efficacy, subsequently updating these evaluations based on treatment outcomes. Moreover, physicians can alternate between the two cognitive systems. To model this behavior, we construct a hybrid cognitive model employing an ε_n-greedy strategy, which increasingly favors rational decision-making as patient outcome data accumulates. Our findings indicate that, when no bias is present, both intuitive and rational systems maintain a nonzero probability of choosing the less effective treatment. When underrepresentation bias in the clinical trials is introduced, we show that the incomplete learning (i.e., the choice of the less effective treatment) resulting from the Gittins index policy in the rational system might be exacerbated. However, our key takeaway shows that the hybrid cognitive model overcomes this bias such that the physician eventually chooses the more effective treatment.

4 - Primary Care Visit Regularity: How and when Does It Matter?

Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States, Vishal Ahuja, Aravind Chandrasekaran

Patients with chronic conditions visit their primary care providers (PCPs) repeatedly. For such patients, regular visits are an important part of care management process. Variations in visit regularity can be detrimental to patients as it could result in untimely or inappropriate care, potentially leading to adverse outcomes. Our research studies this in the context of diabetes. We study how regularity affects health outcomes and the conditions under which such effects are strengthened.

MB56

Summit - 448

Academic and Industry Job Search

Panel Session

Minority Issues Forum

Co-Chair: Pushpendra Singh, Dartmouth College, Hanover, NH, United States

1 - Moderator Panelist

Pushpendra Singh, Dartmouth College, Hanover, NH, United States

2 - Panelist

Sushil Mahavir Varma, Georgia Tech, Atlanta, GA, United States

3 - Panelist

Ying Jin, Data Science Institute, Harvard University, Cambridge, MA, United States

4 - Panelist

Xubo Yue, Northeastern University, Boston, MA, United States

5 - Panelist

Xian Yu, The Ohio State University, Columbus, OH, United States

6 - Panelist

Yeganeh Alimohammadi, Stanford University, Stanford, CA, United States

MB57

Summit - Terrace Suite 1

Advances in Healthcare

Invited Session

Health Applications Society

Chair: Joel Goh, NUS Business School, Singapore, Singapore

Co-Chair: Yvonne Huijun Zhu, National University of Singapore, Singapore, Singapore

1 - Privacy-Preserving Resource Allocation: near Optimal Algorithms and Policy Implications

Chen Du, Nanyang Technological University, Singapore, Singapore, Geoffrey Chua

When multiple agents compete for limited resources, how should the resources be optimally allocated to maximize total utility while preserving agents' privacy, such as utility functions, consumption functions, and special needs? This question sums up today's challenges in balancing the provision of two essential public goods, i.e. healthcare and personal privacy. To address such offline resource allocation problems with privacy concerns, we propose an algorithm called "Noisy Dual Mirror Descent" that privatizes dual variables for privacy and exploits primal-dual schemes for allocation decisions. We show that the algorithm is near optimal and is able to provide a formal privacy guarantee in terms of joint differential privacy. Numerical studies justify our findings and further shed light on policy implications.

2 - A Marginal Distributions Constrained Optimization Approach of Estimating True Prevalence of Pediatric Mental Health Conditions in Washington State

Shan Liu, University of Washington, Seattle, WA, United States, Yinsheng Wang, Kathryn Brandeau

Pediatric mental health disorders such as depression, anxiety, and behavior disorder significantly impact the well-being of children and youth. In Washington state (WA), data has shown that more than one third of people live in areas lacking mental health providers. Accurately mapping the geographic distribution of true demand in WA is crucial for designing an optimal continuum of care network for pediatric mental health services. This study began by estimating the true prevalence of several youth mental health conditions utilizing multiple data sources such as the National Survey of Children's Health (NSCH), WA Healthy Youth Survey (HYS) and American Community Survey (ACS). We developed multivariate regression models using the NSCH data informed by demographic factors, including age, race, sex, insurance status, poverty status, family size, caregiver employment status, etc. To estimate the prevalence of each mental health condition on the census tract level in WA, we constructed a marginal distributions constrained optimization model. The model recovers detailed demographic groups proportions by harmonizing regression results from the NSCH data to estimates in other dataset such as the HYS data, under the constraints of marginal distributions found in the ACS. This approach facilitates the integration of national and local datasets to improve small area estimation and supports the strategic placement of new facilities and programs in downstream research. Overall, this study demonstrates effective data fusion to enhance prediction quality and service planning.

3 - Rl OR Url: Managing Outpatient (Tele)Visits with Strategic Behavior

Nan Liu, Boston College, Chestnut Hill, MA, United States, Shan Wang, Noa Zychlinski

As outpatient care increasingly incorporates telemedicine, providers face the challenge of integrating virtual (URL, or internet) services with traditional in-person (RL, or real-life) services. Whereas telemedicine offers convenience, reducing waiting and travel costs for patients, it may require follow-up in-person visits, which could offset initial benefits. This study examines how providers can optimally balance virtual and in-person service delivery, and explores interventions that can enhance virtual service effectiveness and reduce the need for follow-up in-person visits.

4 - Optimizing Nurse Staffing: Balancing Resource Pooling and Surge Capacity in Health Networks

Zhuoyang Liu, Stanford University, Stanford, CA, United States, Yue Hu, Yueyang Zhong

We explore the joint impact of two hedging methods against demand uncertainty—resource pooling and surge capacity—on nurse staffing in a multi-unit health network. Pooled nurses are pre-staffed and allocated swiftly, while surge nurses are real-time staffed at a higher cost. We assess them based on staffing costs and demand corrections.

MB58

Summit - Terrace Suite 2

Careers in Healthcare Panel Session

Panel Session

Health Applications Society

Co-Chair: Yinsheng Wang, University of Washington, Seattle, WA, 98195, United States

Co-Chair: Meghan Meredith, Georgia Institute of Technology, Atlanta, United States

1 - Moderator Panelist

Meghan Meredith, Georgia Institute of Technology, Atlanta, GA, United States

2 - Moderator Panelist

Yinsheng Wang, University of Washington, Seattle, WA, United States

3 - Panelist

Natalia Summerville, Memorial Sloan Kettering Cancer Center, Cary, NC, United States

4 - Panelist

Lance Waller, Emory University, Atlanta, GA, United States

5 - Panelist

Reza Yaesoubi, Yale School of Public Health, New Haven, CT, United States

6 - Panelist

Md Hafizul Islam, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States

MB59

Summit - Ballroom 1

APS Distinguished Speaker: Jim Dai

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - The BAR-approach for Continuous-time Stochastic Systems

Jim Dai, Cornell University & CUHK-Shenzhen, Ithaca, NY, United States

I will give a tutorial on the basic adjoint relationship (BAR)-approach for steady-state analysis of continuous-time stochastic systems. In such a system, the distributions of stochastic primitives (e.g, service times) are assumed to be general, not necessarily exponential or phase-type. I will illustrate that the BAR-approach is an effective tool to prove (1) the state space collapse in a join-shortest-queue system, (2) asymptotic steady-state independence for queueing networks in multi-scale heavy traffic.

MB60

Summit - Ballroom 2

Amazon Supply Chain Optimization

Invited Session

The Practice Section of INFORMS

Chair: Garrett van Ryzin, Amazon, New York, NY, United States

1 - Dynamic Package Route Computation at Scale

Luciana Buriol, Amazon, Bellevue, WA, United States, Yuri Shevchenko, Olivier Durand de Gevigney, Nithin Lingala, Thomas Helleboid

The journey of a customer package through the Amazon outbound network starts at an inventory location and can go through other Amazon buildings (Sortation Centers, Delivery Stations) and ground, air, rail, and sea legs. Package routing must be compliant with customs and other transportation restrictions. Finally, part of the Amazon demand is delivered by 3P carriers. To delight our customers the routes need to be fast and cheap. In this talk we will describe the first stage of this journey: determining the package routes. For that we developed an algorithm engineering framework that enables fast and dynamic route computation at scale.

2 - Consensus Planning for Capacity Planning: A Market-based Approach to Coordination at Amazon

Garrett van Ryzin, Amazon, Brooklyn, NY, United States

Managing Amazon's fulfillment network is extraordinarily complex due to its vast scale and scope and optimizing overall supply chain performance requires a high degree of coordination across multiple organizations and systems. To address this problem, we developed a distributed, market-based approach to coordination called Consensus Planning Protocol (CPP). In this talk, we describe CPP and how it is being used at Amazon to build an automated planning system (APS) for supply chain capacity planning. APS coordinates complex capacity decisions that were previously made separately and uses scenario-based optimization to jointly hedge capacity decisions against supply and demand risk. The modular, agent-based design enables us to build APS quickly, and easily scale and expand the system in the future.

3 - Consensus Planning with Primal, Dual, and Proximal Agents

Alvaro Maggiar, Amazon.com, New York, NY, United States

The Consensus Planning Protocol (CPP) is a coordination mechanism to align different distributed agents (systems) that share common resources and who act in order to optimize a joint goal. The framework has been adopted within and outside of SCOT to coordinate complex systems. Nonetheless, the interfaces with the agents can be of different types, primal, dual, and proximal, and most of the work thus far has restricted agents to have the same interface, or be either primal or dual. We present in this work a general CPP algorithm that allows for any mix of agents. We prove convergence results for the algorithm, namely a sublinear $O(1/k)$ convergence rate under mild assumptions, and two-step linear convergence under stronger assumptions. We additionally discuss enhancements such as acceleration and provide illustrating results.

4 - A primal recovery method for improving convergence speed of distributed algorithms

Tetiana Parshakova, Amazon, New Haven, CT, United States

We address the problem of minimizing a function that is a sum of agent functions subject to affine constraints that couple them. We assume the agent functions can only be accessed via a conjugate subgradient oracle, meaning agents return their optimal actions given prices. We introduce a new primal recovery method that uses a convex combination of epsilon-optimal solutions to reduce the optimality condition residuals. In the nonconvex case, the produced new primal variable approximately solves a nonconvex problem. Through various examples we show that our method typically converges to a reasonable approximate solution in a modest number of iterations.

MB61

Summit - Ballroom 3

What’s in a Name?

Panel Session

Committee’s Choice

Co-Chair: Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

1 - Moderator Panelist

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

2 - Panelist

Michael Trick, Carnegie Mellon University in Qatar, Doha, Qatar

3 - Panelist

Jeffrey Camm, Wake Forest University, Winston-Salem, NC, United States

4 - Panelist

Sheldon Jacobson, University of Illinois, Urbana, IL, United States

MB62

Summit - Signature Room

Combining Large Language Models and OR/MS to Make Smarter Decisions

Invited Session

TutORial

Chair: Harish Krishnan, University of British Columbia, Vancouver, BC, Canada

1 - Combining Large Language Models and OR/MS to Make Smarter Decisions

Segev Wasserkrug, IBM Research, Haifa, Israel, Leonard Boussioux, Wei Sun

Operations Research/Management Science (OR/MS) capabilities can provide tremendous value in helping enterprises and individuals make smarter decisions. However, the creation and deployment of OR/MS-based decision-making solutions require significant time and expertise, making their widespread application challenging. LLMs, exemplified by models such as ChatGPT, Gemini, and Claude, are deep neural network models encompassing billions/trillions of parameters. These models are pre-trained with a vast scope of general knowledge and quickly adaptable to many downstream tasks. Beyond core capabilities LLMs exhibit emerging capabilities, such as learning new tasks from a few natural language examples. Generative AI technologies like LLMs, have transformative potential for many fields. In this work, we explore the potential of LLMs and their capabilities to significantly improve the creation of OR/MS based decision-making solutions. We show this potential through three concrete and detailed examples: using LLMs to significantly reduce time required to create decision-making applications while improving their quality; using LLMs to extract structured information from unstructured text without the need to create new natural language processing models, for demand forecasting; and using LLMs to drive natural language-based interfaces enabling a business user to easily and flexibly interact with analytical models used for decision making. As LLMs are a new technology that introduces new risks, this work also provides guidelines for their productive, ethical and responsible use and describes ongoing developments relevant to the OR/MS profession. We hope this work will serve as a starting point both for the application of LLMs to the OR/MS use cases we described.

MB63

Regency - 601

Navigating the New Frontier: The Intersection of Artificial Intelligence and the Future of Work

Invited Session

Information Systems

Chair: J. Frank Li, UBC Sauder School of Business, Vancouver, BC, Canada

1 - Knowing More is Better? Examining the Influence of Customer Knowledge Transfer on Customer Service Outcomes

Zixi Lei, University of Texas at Austin, Austin, TX, United States, Wen Wen, Cheng Zhang, Andrew Whinston

This study explores the impacts of customer knowledge transfer among workers on service outcomes in service-oriented industries. Our empirical analysis reveals that customer knowledge transfer among workers significantly improves service outcomes, measured by customer spending. By categorizing customer knowledge into personality, family, service preferences, service descriptions, and recommendations, we also analyze how different types of customer knowledge influence service outcomes. We find that while knowledge about personality traits and service recommendations significantly boost customer spending, the use of sensitive customer knowledge such as their family can negatively impact spending, especially among privacy-sensitive customers. Our findings highlight the dual nature of customer knowledge transfer: it can enhance service outcomes when managed carefully but may backfire if perceived as intrusive. This research contributes to the literature by bridging the gap between customer data use and intra-organizational knowledge transfer. It also provides practical implications for effectively leveraging customer knowledge in organizational settings.

2 - "Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets

Xi Nan, University of Washington, Seattle, WA, United States, Jin Liu, Xingchen Xu, Yongjun Li, Yong Tan

Large Language Model (LLM) based generative AI, such as ChatGPT, is considered the first generation of Artificial General Intelligence (AGI), exhibiting zero-shot learning abilities for a wide variety of downstream tasks. Due to its general-purpose and emergent nature, its impact on labor dynamics becomes complex and difficult to anticipate. Leveraging an extensive dataset from a prominent online labor market, we uncover a substantial post-ChatGPT decline in labor supply, demand, and transactions for submarkets pertaining to text-related and programming-related job categories, in comparison to those not directly exposed to ChatGPT's core functionalities. Meanwhile, these affected submarkets exhibit a discernible increase in the complexity of the remaining jobs and a heightened level of competition among freelancers. Intriguingly, we find that the decrease in programming-related labor supply is relatively smaller, which is attributed to the increasing participation of freelancers who previously engaged in text-only jobs. Although the overall job diversity freelancers bid for per period tends to be more limited, those who successfully navigate skill transitions from text to programming demonstrate greater resilience to ChatGPT's overall market contraction impact. As AI becomes increasingly versatile and potent, our paper offers crucial insights into AI's influence on labor markets, the gig economy, and pertinent regulations, underscoring the necessity for proactive interventions to address the challenges and opportunities presented by this transformative technology.

3 - Unpacking the AI Blackbox: The Impact of AI Strategies on Firm Performance with a Dual Lens on Product and Process Orientation

Jaecheol Park, University of British Columbia, Vancouver, BC, Canada, myunghwan lee, J. Frank Li, Gene Moo Lee

Artificial intelligence (AI) technologies have become increasingly pervasive and hold great potential for large-scale economic impact. Aligned with this trend, recent studies explore the adoption and use of AI technologies on firm performance. However, they predominantly focus on AI as an input (e.g., labor/job posting), neglecting to consider the strategies to leverage their AI inputs in business operations. Thus, it is crucial to understand "how" and "where" to use AI to achieve business value. In this paper, we examine how firms' strategic AI orientation affects firm performance with a dual lens on product and process orientation. We measure AI strategic orientation by employing LLM to assess it from business descriptions in Form 10-K filings between 2015 and 2022. Our findings show that 13% of firms have AI orientation, 7% have AI product orientation, and 3% have AI process orientation. Additionally, we will provide some preliminary results on the impact of AI orientation on firm performance.

4 - The Impacts of Generative AI on Software Development

Sardar Fatooreh Bonabi, University of California, Irvine (UCI), Irvine, CA, United States, Sarah Bana, Vijay Gurbaxani, Tingting Nian

Large Language Models (LLMs) like ChatGPT, have significantly influenced various occupational fields. This study investigates their impact on the software development sector, focusing on productivity and knowledge dissemination. Leveraging a natural experiment resulting from a temporary ChatGPT ban in Italy, our study utilizes a Difference-in-Differences approach to assess the impact of ChatGPT on software development. We compare activity data from GitHub users in Italy with users in France and Portugal to isolate the effects of ChatGPT on productivity and knowledge dissemination within the sector. Our analysis covers a 16-week period, including pre-ban, during-ban, and post-ban phases, involving 95,164 users.

Our findings reveal that access to LLMs enhances software development productivity outcomes by an average of 5.4%. This may be because LLMs enable developers to complete their tasks more rapidly and effectively. More intriguingly, the lifting of the ban led to a 7.8% increase in knowledge dissemination activities, such as reviewing pull requests and opening issues, highlighting how the saved time through LLM usage is redirected towards collaborative and knowledge-sharing efforts. These results suggest that employing LLMs not only boosts productivity but also enhances collaborative learning and knowledge exchange within the software development community.

This paper contributes to understanding the broader implications of AI in the workplace, emphasizing how AI tools can transform not just individual productivity but also collective industry dynamics. Our study underscores the potential of generative AI to advance industry standards, efficiency, and innovation, suggesting a future where AI tools are integral to the software development ecosystem.

MB64

Regency - 602

Modern Statistics in Social Media Analysis

Invited Session

Social Media Analytics

Chair: Alexander Gilgur, Amazon Kuiper, Sunnyvale, CA, United States

1 - Inferring Unusual Metrics from Social Media

Alexander Gilgur, Amazon Kuiper and Stevens Institute of Technology, Sunnyvale, CA, United States

Social media is a trove of information, open for mining without violating people's privacy. With some creativity, statistics-based machine learning tools can be applied off the shelf and / or created new to infer unhappiness, resilience, and even understand the importance of explanatory variables. In this presentation, the author shares some of these ideas that were developed over the course of 6 years of research.

2 - Causal inference through privacy compliant experimentation that makes social advertising more relevant

Carlos Avello, Amazon, Seattle, WA, United States

Conventional Machine Learned marketing operations run on troves of observational data (impressions, clicks, conversions). Observational data is marred by confounders (treatment and effect being correlated beyond causality) which leads to unreliable indicators of incremental effects of the marketing treatment that we run our ML marketing operations for. Therefore, in order to optimize to true incremental effect, ground truth causality information needs to be integrated into our systems. We will describe how to leverage highest science standard of causality via Randomized Control Trials in Social advertising while complying with latest privacy guarantees (DMA). Then we will describe ML marketing operation systems that incorporate RCT data as feedback loop and how it improves relevance and hence greater incremental return to marketing spend

3 - Enhancing low-resource LLMs classification with PEFT and synthetic data

Parth Patwa, Amazon, Santa Clara, CA, United States

Large Language Models (LLMs) operating in 0-shot or few-shot settings achieve competitive results in Text Classification tasks. In-Context Learning (ICL) typically achieves better accuracy than the 0-shot setting, but it pays in terms of efficiency, due to the longer input prompt. In this talk, we describe a proposed a strategy to make LLMs as efficient as 0-shot text classifiers, while getting comparable or better accuracy than ICL. Our solution targets the low resource setting, i.e., when only 4 examples per class are available. Using a single LLM and few-shot real data we perform a sequence of generation, filtering and Parameter-Efficient Fine-Tuning steps to create a robust and efficient classifier. Experimental results show that our approach leads to competitive results on multiple text classification datasets.

4 - Engagement Prediction in Short Video Sharing Platforms —The Case of a Health Campaign on TikTok

Amirmohammad Paksaz, University of Iowa, Iowa City, IA, United States, Kang Zhao, Bingbing Zhang

In recent years, short video sharing platforms, such as TikTok, have emerged as powerful platforms for raising public's awareness about health-related issues through launching health awareness campaigns such as #BreastCancerAwareness. While initial research has begun to explore the features of health-related campaign videos on such platforms, limited research has conducted large-scale examinations beyond a couple of hundred videos. At the same time, understanding the characteristics of health campaign videos systematically and predicting users' engagement with such videos, particularly those related to #BreastCancerAwareness, is crucial due to their demonstrated positive impact on community building and social support. It is also important for video creators to know if their videos will be engaging and, if not, how they can better design their videos. To address this gap, we collected a large dataset of health campaign videos related to #breastcancerawareness during breast cancer awareness month (October 1st, 2023-October 31st, 2023) from TikTok, the most popular short video sharing platform. By examining characteristics of these videos (e.g., length, resolution, and content) and their publishers (e.g., behavior history and account descriptions), we will build multi-modal machine learning models that predict how users engage with these videos. Such a prediction will be the foundation of more in-depth analyses on how to run a more effective health campaign on short video sharing platforms.

MB65

Regency - 603

Power of LLMs in the Analytics Classroom: The Good, the Bad, and the Inevitable II

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Sourav Chatterjee, University of North Texas, 1155 Union Circle, #311160, Denton, TX, 76203, United States

1 - Analytics in the Age of Generative AI

Sourav Chatterjee, University of North Texas, Denton, TX, United States

In this study, we explore the adoption, use, and impact of Generative AI within the business analytics field. Specifically, we investigate how LLMs are transforming the data analytics workflow. Instead of focusing on manual tasks like data cleaning, model building, and coding, business analytics professionals now need to spend more time assessing and managing analyses performed by the LLM models. To understand the impact of this transition, we interviewed industry leaders across various sectors. Our findings highlight an increased emphasis on problem-solving, critical thinking, and domain knowledge. Additionally, the shift toward LLM-driven analytics pushes for a more interactive and personalized learning environment for students pursuing an analytics degree.

2 - Closing the Gap: Leveraging GenAI for Inclusive Data Learning

Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, PA, United States

This session explores the integration of GenAI tools in the classroom to enhance student comprehension and proficiency in exploratory data analysis and visualization. Through hands-on exercises and group activities, students understand the impact of GenAI, enhance their data literacy, boost their confidence in coding and become more engaged in the learning process. We address concerns regarding over-reliance on GenAI tools by advocating for a balanced approach that emphasizes critical thinking and understanding of coding principles, ultimately preparing students for a life-long journey of tech learning.

3 - Using Gen AI in an Undergraduate Business Analytics Course: An Informal Case Study

Manoj Chari, University of North Carolina at Chapel Hill , Chapel Hill, NC, United States

We will discuss some experiences of incorporating Gen AI into a Python based undergraduate Business Analytics course as an assistive tool to enhance student learning of basic python programming for analytics. The goal was to allow and, in fact, encourage students to use Gen AI for feedback and guidance for programming homework and practice work, thus augmenting personal individual guidance available during the class and office hours. We will also discuss the supporting material that was created, the guidelines the students were expected to follow on the use of Gen AI, the adaptation of the course assessment in this context, and anecdotal evidence of students' actual use of Gen AI and students' perceptions of its learning benefits.

4 - Generative AI & LLMsfor Analytics: Fairness-informed Pedagogy

Deepak Khazanchi, University of Nebraska Omaha, Omaha, NE, United States

In AI & LLM applications, unfairness is triggered due to bias in the data, curation process, erroneous assumptions, and implicit bias rendered within the ML development process. As ML applications come into broader use in teaching analytics, it is important the students and instructors understand the attributes necessary for developing fair AI & LLM applications. We have proposed and *prima facie* validated a multidimensional notion of ***perception of fairness*** that should be ideally integrated as a means to assure the use of AI & LLMs in the analytics teaching enterprise.

MB66

Regency - 604

AI in SCM

Invited Session

Artificial Intelligence

Chair: Huihui Chi, NEOMA business school, paris, France

Co-Chair: Michael Lash, University of Kansas, Lawrence, KS, United States

1 - Post-Disaster Emergency Supply Chain Optimization: Multi-Period and Multi-Variety Supplies Allocation and Distribution Amid Uncertain Demand

Shimin Gao, Central University of Finance and Economics, Beijing, China, People's Republic of, Hui Li, Xi Wang, Jian Wang

Disasters like earthquake, flood and snowstorm not only threaten lives of human beings, but also make a huge impact on social and economic development. After the disaster, the damage to communication and road capacity could cause an explosive growth in demand for emergency supplies, and the demands for different types of emergency supplies may vary with the period. Due to the bad communication between disaster area and outside, the pattern of demands generation also shows a strong randomness. How to optimize the routing problem in emergency supply chain, find an effective and robust strategy among stochastic requests, became the popular concern in the research filed of SCM.

We address this challenge by presenting a novel anticipatory routing, acceptance and postponement policy for the Multi-period Dynamic Vehicle Routing Problem with Stochastic Requests (MDVRPSR). We first show an example that demonstrates the features of MDVRPSR well. Then, we apply a Markov Decision Process to establish the mathematical model of this problem. Considering the problem's features, our model is designed to response stochastic requests at multi-period dynamic procedure, in which gives a detailed perspective of the decision state, reward setting, transition procedure and objective function. After that, to acquire the effective policy, we use deep reinforcement learning method like Deep Q-Network to solve this dynamic problem. Computational experiments are conducted to find possible optimal policies in this problem. Finally, we give a full review of our work, evaluate the effectiveness and applicability of our policy.

2 - Does Selling More Expiring Food Reduce Waste? a Perspective of Retail Inventory Allocation

Huihui Chi, EM Lyon, Écully, France

Emerging sustainable logistics systems for perishable products (such as food and pharmaceutical products) have been proposed in the recent past with the aim of reducing food waste and improving food quality. Trying to answer the practitioners' question of "does selling more expiring milk reduce milk waste?" from the inventory allocation perspective, in this research, we develop a two-stage model with shelf-space-dependent demand, where the retail shelf space is limited for placing two perishable products with long (high quality) and short (low quality) remaining shelf life. We identify the optimal allocation preferences between the capacitated shelf space and uncapacitated warehouse with stochastic demand. We calculate the bounds on the optimal solution that are integrated in the heuristic method. In conclusion, we discuss the managerial implications of food waste reduction from the perspective of inventory allocation strategy.

3 - Post-hoc Interpretability using Counterfactually-guided Multi-agent Contextual Bandits

Michael Lash, University of Kansas, Lawrence, KS, United States

Machine learning (ML) explainability (MLX) allows decision-makers to assess whether an ML model is making predictions "for the right reasons" and is thus crucial to the AI-powered decision-making pipeline. Obtaining faithful, quality explanations is therefore equally crucial. To such an end, this work proposes a multi-agent counterfactually guided contextual bandit approach to the MLX problem. The method is designed to synthesize explanation-producing policies using an off-policy counterfactually-guided search scheme. Our experiments show that the explanations obtained using our proposed method are superior to those obtained from competing methods.

4 - Toward Trustworthy AI: Mitigating Misleading Explanation's Risk in Machine Learning Predictions

Amir Asrzad, University of Massachusetts Lowell, Lowell, MA, United States

Machine learning, especially deep learning, excels at tackling complex prediction tasks with impressive accuracy. However, the lack of transparency in these "black-box" models raises concerns. Explainable AI (XAI) techniques aim to bridge this gap by providing human-interpretable explanations for model decisions. Yet, XAI explanations can sometimes be insufficient or misleading, leading to user distrust and hindering the adoption of AI technology. Despite its significance, the XAI literature lacks robust strategies to mitigate the risk of misleading explanations. This study addresses this gap by introducing a novel approach for generating risk-sensitive counterfactual explanations. Our method balances two key challenges: Overly weak explanations: We create robust counterfactuals to avoid explanations lacking reliable insights. Uninformative explanations: We introduce alert counterfactuals to prevent misleadingly late suggestions. The effectiveness of our proposed method is rigorously evaluated on real-world datasets to demonstrate its empirical value.

MB67

Regency - 605

2024 INFORMS Prize

Invited Session

INFORMS Prize 2024

Co-Chair: Yulia Vorotyntseva, University of St. Thomas, Minneapolis, MN, United States

1 - 2024 INFORMS Prize Winner JD.com

Hau Lee, Stanford University, Stanford, CA, United States

JD.com, also known as Jingdong,marked its entry into the e-commerce sector in 2004 and is a leading supply chain-based technology and service provider. The company's expansive portfolio spans across sectors including retail,technology, logistics, health, and more,underscoring its diverse and multifaceted operations. In 2023, JD.com ranked 52nd on the Fortune Global 500 list.The company serves nearly 600 million active users, supported by an extensive infrastructure, ensuring unparalleled service and efficiency in its operations.JD. com has long been committed to applying operations research and management science to solve practical problems and has been selected as a finalist for the Franz Edelman Award twice,in 2021 and 2023. JD. com has been named the 2024 INFORMS Prize recipient its outstanding utilization of operations research and data science methodologies,driving innovation and excellence across its supply chain and e-commerce operations.

MB68

Regency - 606

Analytics in Healthcare Decision Making

Invited Session

Analytics Society

Chair: Charles Nicholson, The University of Oklahoma, Norman, OK, United States

1 - Scalable and Trustworthy Deep Neural Networks for Imbalanced Data

Rachel Bennett, University of Oklahoma, Oklahoma City, OK, United States, Talayeh Razzaghi

Advancements in data availability have significantly placed neural networks at the forefront of successful learning algorithms, particularly for healthcare applications. However, training with large datasets requires extensive resources, and presents notable challenges, such as managing class imbalance and ensuring fairness. To address these challenges, we introduce a novel Multilevel Deep Neural Network (MLDNN) approach, which is designed to scale effectively with large datasets and provide a robust solution to class imbalance. We will demonstrate that MLDNN offers more accurate and robust classification results compared to traditional models using both healthcare-specific and benchmark datasets.

2 - A Systematic Review of Stochastic Optimization Techniques in Healthcare Decision Making: Methods, Applications, and Future Directions

Mansura Roly, The University of Oklahoma, Norman, OK, United States, Charles Nicholson

Stochastic optimization techniques have emerged as powerful tools for decision support in healthcare systems, offering robust methodologies to address the uncertainties inherent in medical decision-making processes. This presentation synthesizes recent advancements in the application of stochastic optimization models within the healthcare domain, focusing particularly on their utilization in decision-making scenarios such as resource allocation, treatment planning, and operational management. Through a comprehensive survey of the literature, this presentation explores the various stochastic optimization methodologies employed, including stochastic programming, Markov decision processes, and simulation optimization, among others. By providing a systematic overview of the existing research landscape, this work aims to inform healthcare practitioners, policymakers, and researchers about the potential benefits and limitations of stochastic optimization approaches.

3 - Optimal Grid Placement for Treatment Planning of Grid Therapy

M. Soheil Hemmati, University of Oklahoma, Norman, OK, United States, Grant M. Benson, Raghavendiran Boopathy

As a special form of radiation therapy, grid therapy is often used to treat bulky tumors, possibly with irregular shapes. At current, treatment planning is essentially performed based on institutional guidelines for target contouring, which requires placing uniform-size grids within a certain distance from each other, while ensuring the grids are sufficiently spared from the surrounding organs-at-risk (OARs). Further, the process is performed manually with no account for optimality in placing the grids with respect to objectives such as maximum coverage, maximum tumor control, or minimum normal tissue complication probability. We propose a family of mixed-integer programs to formulate the optimal grid placement problem in grid therapy. We solve instances of our proposed models for a set of cancer patients treated with grid therapy at the University of Oklahoma's Health Science Center and compare the outcomes of our models with the existing clinical procedures for treatment planning of grid therapy.

4 - Revisiting the Gateway Hypothesis by Considering the Effect of Age-of-First use on Subsequent Illicit Drug use

Matthew Beattie, University of Oklahoma, Norman, OK, United States

The Gateway Hypothesis claims that drug use evolves in stages. It begins with either tobacco or alcohol use, progresses to use of the other, then to marijuana, then to other more dangerous and illicit substances. Drug use in one stage is required for, but not a determinant of, use in subsequent stages. This study builds upon this established and validated work by showing the relationship between age of first use (AFU) of quasi-licit drugs and the likelihood of use of illicit drugs. Two explainable machine learning techniques, decision tree (DT) and logistic regression (LR), are applied to NSDUH 2016-2019 and 2022 data to model the relationship between determinant features (AFU of tobacco, alcohol, and marijuana) and any lifetime use of other illicit drugs. Both models are accurate, with AUROC values of 0.84 and 0.85 respectively. The DT model accurately predicts 84% of illicit drug users and 74% of non-users. The LR model predicts 79% of users and 77% of non-users. Both models find that marijuana AFU is a far more important predictive feature than alcohol AFU or tobacco AFU, and highlight risks of early marijuana use. These findings point to the need to set policies that improve limitations on access of marijuana to underage users.

5 - Effeciency Vs Education: the University of New Mexico&Rsquo;S Journey to Increased Efficiency in the Context of Emergency Medical Resident Education

Matt Frederick, University of New Mexico, Albuquerque, NM, United States

In 2023 UNM's emergency Department's overcrowding was exacerbated by the state mandated closing of “hall beds.” These beds (literally, in a hallway in the emergency department) allowed a higher number of patients to be out of the waiting room and in a more monitored environment but were technically unlicensed. The outcome of increased waiting room patients into such a stressed system inevitably leads to statistically poorer outcomes, and in a tangible sense often leads to people actually dying in the waiting room, only feet from some of the world’s most advanced resuscitation technology.

A vertical care (VC) area was created which allowed a more rapid evaluation and treatment of patients that would previously have been evaluated only in the waiting room. This unburdened the waiting room staff allowing them to be more efficient in their tasks leading to an overall decrease length of stay and better patient care.

However, the changes necessary for the VC implementation led to dramatic challenges within the educational ecosystem. Resident physicians are a key work force for our facility, and there is an inherent tension between educating future doctors and efficiently treating patients.

Our presentation will present this journey. We will cover the data-based decisions and then explore the human and educational factors that ultimately caused a deviation from our initial design. We will also present the outcome of our intervention, which has in a very real sense saved lives and ultimately improved the educational experience for our learners.

MB69

Regency - 607

Digital Strategy

Invited Session

eBusiness

Chair: Haoyuan Liu, Nanyang Technological University, Singapore, N/A, Singapore

1 - Does Political Leaders&Rsquo; Technical Background Affect Firms&Rsquo; Digital Transformation?

Xinjun Yang, Nanyang Technological University, Singapore, Singapore, Haoyuan Liu, John Qi Dong

Digital transformation as firms’ changes of business models with digital technologies have been reshaping the competitive landscape across industries, yet scholarly understanding of its antecedents remains limited. This study addresses this gap from a resource dependence perspective by investigating the impact of technocrats with technical background in provincial governments on firms’ decisions toward digital transformation. Leveraging on the exogenous assignment of provincial government leaders in China, we find that a higher proportion of technocrats in provincial government leaders correlates with increased willingness of firms in the province to pursue digital transformation. This effect is notably more pronounced among private firms compared to state-owned enterprises and foreign firms, possibly due to resource limitations and thus a higher dependence on external institutional relations with governments.

---

2 - The Strategic Secretary Problem

Yufei Zhu, National University of Singapore, Singapore, Singapore, Joel Goh

In the classical secretary problem, secretaries have no decision agency and are simply assumed to arrive in a random order. In this paper we consider a novel variation of the classic problem. Motivated by modern online labor platforms where job-seekers have much more information at their disposal, we assume that secretaries are strategic agents and can choose their arrival time based on a privately revealed quality score. We consider a game between the hiring firm and N strategic secretaries and analyze the structure of the equilibria of this game. This paper tackles the questions: Do strategic secretaries gain more utility than non-strategic ones? Does the firm's ability to select the best secretary improve or deteriorate when secretaries are strategic?

3 - Echoes of Authenticity: Reclaiming Human Sentiment in the LLM Era

Yi Ding, University of Warwick, Conventry, United Kingdom, Yifei Wang, Ashkan Eshghi, Ram Gopal

This paper scrutinizes the unintended consequences of employing large language models (LLMs) like ChatGPT for editing user-generated content, particularly focusing on alterations in sentiment. Through a detailed analysis of a climate change tweet dataset, we uncover that LLM-rephrased tweets tend to display a more neutral sentiment than their original counterparts. By replicating an established study on public opinions regarding climate change, we illustrate how such sentiment alterations can potentially skew the results of research relying on user-generated content. To counteract the biases introduced by LLMs, our research outlines two effective strategies. First, we employ predictive models capable of retroactively identifying the true human sentiment underlying the original communications, utilizing the altered sentiment expressed in LLM-rephrased tweets as a basis. While useful, this approach faces limitations when the origin of the text—whether directly crafted by a human or modified by an LLM—remains uncertain. To address such scenarios where the text's provenance is ambiguous, we develop a second approach based on the fine-tuning of LLMs. This fine-tuning process not only helps in aligning the sentiment of LLM-generated texts more closely with human sentiment but also offers a robust solution to the challenges posed by the indeterminate origins of digital content. This research highlights the impact of LLMs on the linguistic characteristics and sentiment of user-generated content, and more importantly, offers practical solutions to mitigate these biases, thereby ensuring the continued reliability of sentiment analysis in research and policy.

4 - When Should an Online Retailer Reach Out to Consumers after Their Purchase?

Youngsoo Kim, Texas Tech University, Lubbock, TX, United States

This study attempts to answer when a full-line online retailer selling products from diverse categories (i.e., Amazon.com) should reach out to consumers after their purchase. We collected the data showing 361,263 transactions of 10,103 consumers from an online retailer selling diverse products. Our empirical analyses show that consumers buy at the *Amazon.com-style* online retailer highly irregularly due to increased shopping time flexibility. We find that online retailers need to reach out to consumers with marketing activities immediately after their purchase, because the likelihood of another purchase at the retailer continuously drops as inter-purchase time increases (monotonic decreasing hazard rate). We also confirm two other price promotion effects, in addition to the stockpiling effect of price promotion (i.e., cheaper prices induce more purchase). First, consumers who benefit from price promotions tend to adjust the next purchasing timing for another price promotion; consumers become more price-sensitive after experiencing a price promotion. Second, after consumers benefit from a price promotion, they are likely to purchase sooner at the retailer again than the consumers who purchased without a price promotion. This behavior is rationalized on habitual decision of an online retailer (a good impression after benefiting from a price promotion) or the increased loyalty to the online retailer. Our results provide useful bases for differentiation of marketing campaign timing according to purchase situation (e.g., elapsed time since the last purchase) as well as time-variant individual profile.

MB70

Regency - 701

ENRE: Optimization and Learning for Sustainable Grids

Invited Session

ENRE: Electricity

Chair: Tongxin Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

1 - Reliable Machine Learning-Augmented Algorithms for Energy Systems and Sustainability

Nicolas Christianson, California Institute of Technology, Pasadena, CA, United States

Modern AI and ML algorithms hold immense promise for improving performance in decision-making under uncertainty, where traditional algorithms designed for the worst case may be overly conservative; this can potentially be transformative for energy and sustainability applications, where rapid improvements are crucial to facilitate the energy transition and reduce carbon emissions. However, AI and ML lack worst-case guarantees, hindering their deployment to real-world problems where safety and reliability are crucial. In this talk, I will discuss recent work on developing machine learning-augmented algorithms with provable performance guarantees and their application across the domains of energy systems and sustainable computing. In particular, I will focus on the questions of algorithmic upper and lower bounds for integrating ML and classic, worst-case approaches, how to optimally leverage uncertainty in decision-making, and designing risk-averse algorithms in these settings.

2 - Towards Environmentally Equitable AI via Geographical Load Balancing

pengfei Li, University of California, Riverside, Riverside, CA, United States

Fueled by the soaring popularity of large language and foundation models, the accelerated growth of artificial intelligence (AI) models' enormous environmental footprint has come under increased scrutiny. While many approaches have been proposed to make AI more energy-efficient and environmentally friendly, environmental inequity -- the fact that AI's environmental footprint can be disproportionately higher in certain regions than in others -- has emerged, raising social-ecological justice concerns. This paper takes a first step toward addressing AI's environmental inequity by balancing its regional negative environmental impact. Concretely, we focus on the carbon and water footprints of AI model inference and propose equity-aware geographical load balancing (GLB) to explicitly address AI's environmental impacts on the most disadvantaged regions. We run trace-based simulations by considering a set of 10 geographically-distributed data centers that serve inference requests for a large language AI model. The results demonstrate that existing GLB approaches may amplify environmental inequity while our proposed equity-aware GLB can significantly reduce the regional disparity in terms of carbon and water footprints.

3 - Hierarchical Forecasting in the Energy Domain: A State Estimation Approach

Matthew Motoki, University of Washington, Seattle, WA, United States, Jiayi Li, Baosen Zhang

In the energy sector, data often exhibits a natural hierarchical structure; for example, energy consumption data can be organized from device-level to household-level, to substation-level, and so on. Another natural hierarchy occurs when splitting sources of renewable generation into categories such as solar, wind, and hydro. Accurate and coherent forecasts across all levels in the hierarchy is critical for operational efficiency and strategic planning. In this paper, we present a novel hierarchical forecasting approach that generalizes traditional forecast reconciliation methods by formulating the task of creating coherent forecasts as a state estimation problem. Our method not only improves the accuracy at both the user and system levels but also addresses scalability challenges often encountered with existing techniques. We derive the optimal forecasts for the case where the users' noise is correlated and their target variable is a linear function of their input and show through simulation that our approach outperforms conventional methods. Furthermore, our method works with various objective functions and can be used to minimize the error in peak prediction, which historically has been a challenging problem. Finally, we provide empirical results involving nonlinear forecast models, such as neural networks and gradient-boosted decision trees, and introduce a scalable, distributed algorithm whose effectiveness is validated with real-world data.

MB71

Regency - 702

Data Mining on Electronic Health Records: Extracting Key Features and Patterns Using Machine/Deep Learning

Invited Session

Data Mining

Chair: Zhuqi Miao, SUNY-New Paltz, New Paltz, NY, United States

1 - A Novel Hyperparameter Search Approach for Accuracy and Simplicity in Disease Prediction Risk Scoring

Yajun Lu, Jacksonville State University, Jacksonville, AL, United States, Thanh Duong, Zhuqi Miao, Thanh Thieu, Jivan Lamichhane, Abdulaziz Ahmed, Dursun Delen

This study introduces a novel approach to crafting risk scores for disease prediction models based on logistic regression. The objective is to find the optimal number of regression units per risk point, striking a balance between simplicity and accuracy for effective patient risk stratification. The proposed technique employs an adapted line search method and integrates the DeLong test to ensure accuracy comparable to existing approaches. Evaluation is conducted through case studies predicting diabetic retinopathy (DR) and hip fracture readmissions (HFR), involving large patient cohorts. Results show that our risk scores achieve comparable accuracies to existing methods. Notably, our scores exhibit a compact scale, contrasting with the broader ranges of existing methods. The discussion underscores the simplicity and accuracy of our approach, presenting it as a competitive alternative for disease risk prediction. In conclusion, this study offers a generalizable framework for developing precise risk scores with compact scales, addressing the need for effective risk stratification tools in healthcare.

2 - Empowering Disease Prediction: Data Extraction and Transformation from Cerner AWS Ehr Database

Hieu Nghiem, Oklahoma State University, Stillwater, OK, United States

In the era of data-driven healthcare which focuses on better outcomes, predictive analytics emerges as a transformative tool, offering the potential for accurate prediction of diseases. This leads to improvements in patient outcomes, as well as optimization of resource allocation for proper treatment strategies.

However, the efficacy of predictive models depends on the quality, breadth and diversity of the underlying data. Generally, in a data pipeline to prepare input for machine learning / predicitve modeling, ETL (Extract, Transform, Load) process plays a crucial role. ETL serves as the fundamental gateway, enabling the preparation and transformation of raw data into a format suitable for analysis and modeling. By ensuring data quality, consistency, and accessibility, a proper and efficient ETL process sets the foundation for building accurate, reliable, and effective ML models that can generate valuable insights and drive informed decision-making in various domains, especially in healthcare analytics.

This presentation focus on the data extraction and transformation process from raw data in Cerner EHR databases that are hosted on AWS Cloud (Amazon Web Services). The data is established and distributed through Apache Spark, a widely used unified analytics engine for large-scale data processing in the industry. By performing a comprehensive exploration of data wrangling workflows and methodologies through Python, SQL and PySpark API, we elucidate strategies for effective and accurate data orchestration process in which siloed data is extracted from multiple tables, combined, organized and tailored to match the unique demands of EHR data analysis and disease prediction.

3 - Extracting Patient History from Clinical Text: a Comparative Study of Transformer-Based Clinical Language Models

Zhugo Miao, SUNY New Paltz, The State University of New York, New York, NY, United States

History of Present Illness (HPI) and Past, Family and Social History (PFSH) are two crucial sections of a clinical note, providing valuable information that assists physicians in diagnosing and developing treatment plans. Moreover, they enable effective communication and continuity of care among healthcare professionals involved in the patient's care. However, extracting and analyzing patient history data documented in free-text clinical notes constitute a consistent burden to healthcare professionals. To address this challenge, this study developed a transformer-based named entity recognition model, which can identify detailed HPI and PFSH elements from clinical notes. We evaluated the model's performance and conducted error analysis to identify factors impacting the model's accuracy.

MB72

Regency - 703

AI and Analytics for Good

Invited Session

Data Mining

Chair: Joseph Cazier, Arizona State University, Tempe, AZ, United States

1 - Ai for Good! for Sure?

Ryan LaBrie, Seattle Pacific University, Seattle, WA, United States

There has been a recent trend in artificial intelligence (AI) literature focused on AI advancements for “good”. Good for who? Good for the environment, good for humanity, good for the companies (and their shareholders) that create the technology, or perhaps good for all (or at least many of) those groups. This session seeks to examine and discuss some of the pinch points in which AI might not live up to the hype of being “good for all” and investigate what mechanisms can we use to prevent bias, ensure some level of transparency, or at least reproducibility, and ensure responsible AI usage? Are AI audits a viable solution or are there other frameworks or techniques we can draw upon to ensure an AI justice for all and advancement of human flourishing… for all!

2 - Primum Non Nocere: Guiding Principles for Model Development

Hina Arora, Arizona State University, Tempe, AZ, United States

AI models hold tremendous potential for societal benefit. However, without careful design and implementation, they can also lead to substantial harm. This presentation will use case studies from multiple sectors to extract key design principles for developing models that first do no harm.

3 - Succeeding in AI and Analytics

Joseph Cazier, Arizona State University, Tempe, AZ, United States

Reports on research into best practices for leading societal transformation with AI and Analytics.

4 - New INFORMS Resources to Foster Ethical and Successful AI and Analytics

Terry Rawls, Vulcan Academics, Durham, NC, United States

This presentation details updates to the INFORMS Course Leading in Analytics, including the inclusion of the INFORMS Ethics Course and updates for advances in AI.

MB73

Regency - 704

Advances in Healthcare Analytics

Invited Session

Data Mining

Chair: Syed Hasib Akhter Faruqui, Sam Houston State University, Huntsville, TX, United States

Co-Chair: Julian Carvajal Rico, The University of Texas at San Antonio, San Antonio, TX, 78249, United States

1 - Enhancing AI-Assisted Medical Decision-Making: The Vital Role of Uncertainty Quantification

Shengfan Zhang, University of Arkansas, Fayetteville, AR, United States, Maryam Kheirandish

The use of Deep Neural Network (DNN) models in risk-based decision-making has garnered significant attention across fields such as medicine, finance, manufacturing, and quality control. To mitigate risks, it is crucial to assess prediction confidence or uncertainty alongside the algorithm's overall performance. Transitioning from deterministic and probabilistic DNNs to Bayesian deep learning models allows for quantifying prediction uncertainty through second-order prediction distributions. The Shannon entropy and mutual information help quantify and decompose this uncertainty into epistemic and aleatoric components, essential for uncertainty reduction strategies. However, entropy-based measures can be inconsistent, and mutual information may limit the consideration of multiple sources of epistemic uncertainty. Additionally, assuming errors following normal distributions, despite its simplicity, is inadequate for discrete errors. We propose a framework to quantify prediction uncertainty for DNN binary classification models using the variation-ratio measure. This approach supports risk-based decision-making when discrete errors in predictors are present. Our model addresses two sources of epistemic uncertainty: model parameters and errors in predictors with a known finite discrete distribution. This measure helps detect significant contributions of predictor errors to total uncertainty. Applied to a case study predicting tuberculosis treatment outcomes and 10 simulated datasets, our framework identifies risk-sensitive cases prone to prediction alterations due to predictor errors. Even when assuming epistemic uncertainty arises solely from model parameters, our framework demonstrates superior uncertainty awareness compared to the Monte Carlo dropout method.

2 - Medication Reconciliation with Artificial Intelligence and Operations Research Methods

Xinyu Yao, Carnegie Mellon University, Pittsburgh, PA, United States, Amogh Rao, Holly Wiberg, Rema Padman

Medication reconciliation is a vital intervention in healthcare management during care transitions, which aims at identifying the most accurate and complete list of the patient's medications by comparing the physician's admission, transfer, and discharge orders and addressing potential discrepancies. Ensuring alignment of the drug name, dosage, frequency, and route between the patient's existing medication regimen and the newly prescribed medication, medication reconciliation is promising to play a crucial role in enhancing patient safety and minimizing potential medication errors and adverse drug events (ADE) throughout their healthcare journey. Despite ongoing efforts to implement medication reconciliation, challenges persist in realizing its full effectiveness in practice. In this research, we develop an automated medication reconciliation process based on a combination of artificial intelligence (AI) and operations research (OR) algorithms. Our research objective is to significantly enhance the performance of medication reconciliation, thereby contributing to improved patient outcomes and reduced risks associated with medication-related issues.

3 - Abdominal Aortic Aneurysm Rupture Risk Classification Aided by Random Forest Using Biomechanical, Morphological, And Clinical Surrogates

Juan Restrepo, The University of Texas at San Antonio, San Antonio, TX, United States, Satish C. Muluk, Mark Eskandari, Ender Finol

Abdominal Aortic Aneurysms (AAA) ruptures lead to more than 15,000 deaths every year in the US. If a rupture occurs in the interior areas, up to 90% of patients suddenly die within 48 hours if left untreated or improperly treated. Predicting rupture risk is critical for timely surgical intervention and improved patient safety. The current standard of care is determined by a one-size-fits-all, maximum diameter approach. This study introduces an advanced classification framework using Random Forest (RF) to better predict AAA rupture risk, accounting for biomechanical, morphological, and geometric surrogates to translate clinical assessment towards patient-specific approaches. Our findings highlight the potential of integrating patient-specific surrogates into machine learning frameworks to enhance AAA rupture risk assessment. This approach could significantly aid clinicians in making more informed decisions regarding the necessity and timing of interventions, ultimately leading to personalized patient care and better clinical outcomes.

MB74

Regency - 705

Flexibility Analysis for Renewable Dominated Power Systems

Invited Session

ENRE: Electricity

Chair: Jun Wang, The Univeristy of Hong Kong, Pokfulam Road, Hong Kong, Hong Kong, N/A, Hong Kong

Co-Chair: Yue Chen, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, N/A

Co-Chair: Yunhe Hou, The University of Hong Kong, Hong Kong, Hong Kong

1 - Robust generator maintenance for resilient power systems

yang yang, NUS, Singapore, Singapore

Normal operation of a power system requires that alternating current frequency be maintained at a nomq frequency requires adequate inertia and frequency regulation reserve in the power system, which are primarily provided by online generators. In daily operations, generators that are due for preventive maintenance must be taken offline, and thus an improper maintenance schedule could jeopardize frequency security, as exemplified by the recent Texas power blackout. However, this natural nexus between frequency security and preventive maintenance has been overlooked largely in the literature. We fill the gap by developing a generator maintenance scheduling model over an annual planning horizon that incorporates frequency security constraints with hourly fidelity to meet industrial standards. These constraints essentially amount to scheduling adequate inertia and frequency regulation reserve by considering the uncertain power deficiency and inertia that result from intermittent renewable energy. We hedge the uncertainties by employing a robust optimization approach in which historical data are used to construct ambiguity sets. This inevitably results in an ultra-large-scale robust model because of the hourly fidelity. We show that it can be reformulated as a large-scale mixed-integer linear program. An algorithm based on the progressive hedging idea is then proposed to decompose the model into subproblems that can be solved in parallel. To speed the large-scale optimization, we develop a duality-free cutting-plane method for the subproblems and a novel lower bound for the model to collectively enhance computational efficiency. Compared with the standard progressive hedging algorithm and an L-shaped algorithm with strengthened Benders cuts, our algorithm can be 15 times faster and avoids the out-of-memory problems that arise with these benchmarks.

2 - Adaptive Flexibility Planning for Renewable-Power Charging Stations Under Decision-Dependent EV Diffusion Uncertainty

Yujia Li, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, Feng Qiu, Yixuan Chen, Yunhe Hou

When deploying fast charging stations (FCSs) to support long-distance trips of electric vehicles (EVs), there exist indirect network effects: while the gradual diffusion of EVs directly influences the timing and capacities of FCS allocation, the decisions for FCS allocations, in turn, impact the drivers' willingness to adopt EVs. This interplay, if neglected, can result in uncovered EVs and security issues and even hinder the effective diffusion of EVs. In this paper, we explicitly incorporate this interdependence by quantifying EV adoption rates as decision-dependent uncertainties (DDUs) using decision-dependent ambiguity sets (DDASs). Then, a two-stage decision-dependent distributionally robust FCS planning (DR-FCSP) model is developed for adaptively deploying FCSs with on-site sources and expanding the coupled distribution network. A multi-period capacitated arc cover-path cover (MCACPC) model is incorporated to capture the EVs' recharging patterns to ensure the feasibility of FCS locations and capacities. To resolve the nonlinearity and nonconvexity, the DR-FCSP model is equivalently reformulated into a single-level mixed-integer linear programming by exploiting its strong duality and applying the McCormick envelope. Finally, case studies highlight the superior out-of-sample performances of our model in terms of security and cost-efficiency. Furthermore, the byproduct of accelerated EV adoption through an implicit positive feedback loop is highlighted.

3 - Flexiramp in Real-time Balancing: Product Introduction, Bidding Strategy Analysis, and Procurement Mechanism Optimization

Ziqing Zhu, Hong Kong Polytechnic University, Kowloon, Hong Kong

Flexiramp products are widely adopted in the United States electricity markets, such as CAISO and MISO, to address load uncertainty during real-time operations by incentivizing generators to reserve flexible ramping capacity within specific time intervals. However, the procurement and pricing mechanisms of this product remain open research questions, lacking simulation-based insights into generator bidding behaviors and social welfare under different mechanisms. This talk will first introduce existing pricing and procurement mechanisms for flexiramp products, followed by an exploration of utilizing multi-agent reinforcement learning methods to simulate bidding games in electricity markets with flexiramp products. Finally, we will analyze simulation results to derive optimization strategies for pricing mechanisms.

4 - Advanced Optimization for transportation electrification

Lusha Wang, University of Alabama, Tuscaloosa, AL, United States

Transportation Electrification will significantly reduce carbon emissions and reliance on fossil fuel. The world has seen an accelerating transition to electric vehicles (EVs) driven by environmental concerns, government promotions and battery technology advancements. On one hand, this transition brings in great positive environmental impacts; on the other hand, it introduces enormous challenges to existing power systems which were designed and constructed without considering massive EV integration. It is thus imperative to rethink the planning and operations of power grids and pave the way for EV everywhere. This talk will introduce challenges and opportunities associated with transportation electrification, especially in power distribution systems. In terms of operation, new optimization and economic models will be discussed to unleash EV flexibilities and enable various vehicles-to-grid (V2G) services such as voltage control, power regulation and outage restoration; in terms of planning, we will introduce new models for charging station placement considering both electricity and transportation system constraints and their interactions. These proposed methods can help stakeholders including governments, utilities, third-party operators and consumers, to mitigate EV impacts on power systems, leverage EVs to improve grid operations and renewable energy integration, as well as make optimal infrastructure investments to support a sustainable and equitable EV transition.

MB75

Regency - 706

Futuristic Planning in Energy Generation and Markets

Invited Session

ENRE: Environment and Sustainability

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, 80401, United States

1 - Enhancing An Industrial-Scale Linear Program to Plan Capacity for a Power Generation System

Kathleen Tomon, Colorado School of Mines, Golden, CO, United States, John Ayaburi, Edikan Udofia, Alexandra Newman

The Regional Energy Deployment System (ReEDS) is a capacity planning model to examine the long-term effects of technology changes, market evolution, and policy decisions. ReEDS minimizes total capital and operational cost while adhering to capacities, demands, policies and technology interaction constraints. Because of added model complexities, solution times have grown. Previous work on improving the formulation led to a 35% reduction in model run time. We seek to further decrease solve times by enhancing formulation efficiency and employing advanced solution techniques.

This work is sponsored by GAMs.

2 - Optimizing Design and Dispatch of a Microgrid with Heterogeneous Heating Loads

Alexander Zolan, National Renewable Energy Laboratory, Golden, CO, United States, William Becker, An Pham

The adoption of distributed energy resources to serve all or part of a building's electrical load is growing in response to the declining cost of onsite renewables and energy storage. In addition to this, electrification of the built environment motivates the inclusion of thermal loads when considering technology selection to meet a building's combined energy needs, especially when considering decarbonization goals. We introduce REOpt, an open-source web tool that obtains cost-minimizing distributed energy resource designs subject to meeting electrical, heating, and cooling loads with or without a utility connection for residential and commercial customers. We present an extension to include multiple heating loads of different quality and demonstrate the impact of this extension on design selection. We then present case studies showing the impact of technology availability, costs, and emissions reduction goals on technology selection.

3 - Evaluating the State of Optimization Techniques in Environmental Sustainability Metrics for Mine Planning

Raymond Kudzawu-D'pherdd, Colorado School of Mines, Golden, CO, United States, Kwame Awuah-Offei, Esteban Koberg de La Cruz, Marcos Goycoolea, Alexandra Newman

Environmental metrics, critical for evaluating mining impacts in the 21st Century, are inconsistently applied in the mine planning processes. This mine-specific arbitrariness often leads to overlooked environmental considerations at the mine planning stages. Community opposition, driven by environmental concerns and a general distrust towards mining regulatory authorities and mining companies, has been amply documented from across the globe. In response, the concept of "sustainable mining" has emerged, aiming to harmonize community interests and mitigate opposition by focusing on environmental sustainability in mining operations. Despite the growing acknowledgment of the benefits of sustainable mining, there is a noticeable lack of literature on the effectiveness of current environmental metrics in fostering sustainable outcomes. We seek to address this gap by reviewing the status quo to identify, assess and proffer effective environmental metrics for sustainable mining. Our objective is to provide mining companies with recommendations for integrating environmental sustainability metrics into strategic planning, thereby improving operational efficiency, reputation, and stakeholder relationships. Surreptitiously, we are discovering how these metrics can be included at different stages of the mine planning process for entire mining value chain optimization.

4 - Trade-Offs from Vertical Integration Between Natural-Gas and Electricity Markets

Ramteen Sioshansi, Carnegie Mellon University, Pittsburgh, PA, United States

Electricity systems are becoming more dependent upon natural gas as an electricity-generation fuel. As such, electricity and natural-gas markets are becoming more interconnected. Some electricity and natural-gas markets are vertically integrating, through the merger of electricity and natural-gas suppliers. The market-efficiency impacts of this vertical integration is unclear. To study this question, we develop a Nash-Cournot model of the two interconnected markets. The model is converted into a linear complementarity problem, which allows deriving Nash equilibria readily. The model is applied to a stylized example with a range of parameter values. We find that integration is social-welfare enhancing--which implies that mitigating double marginalization outweighs the exercise of market power. In most cases, the effects of merger can give rise to a prisoner's-dilemma-type outcome, whereby firms have strong incentives to merge but merger is detrimental to producers.

5 - Sectoral Decomposition Methods for Multi-Sector Energy Capacity Expansion Models

Frederico Parolin, Massachusetts Institute of Technology, Boston, MA, United States, Ruaridh Macdonald, Yu Weng

Decarbonizing our energy systems will be significantly less expensive if investments and operations are coordinated across sectors, such as electricity, hydrogen, liquid fuels, etc. This coordination is especially important as technology and policy decisions in one sector will have knock-on effects elsewhere. This has created a need for multi-sector energy capacity expansion models which can co-optimize investments and operations across several sectors, long timeseries and multiple regions. However, adding additional sectors greatly increases the number of variables and constraints in the model, limiting optimizations to a few regions or short representative periods. This risks biasing the results. Decomposition methods offer a means of overcoming this problem, breaking up the original large optimization into an iterative process between several smaller problems. Bender's decomposition has successfully been applied to decompose models in time and in this work we have extended that method to also decompose models across sectors. In this presentation we describe the method and demonstrate the performance and accuracy compared to alternative approaches for a large model of the US lower-48 states. We show the reduction in runtime and memory usage versus model scale and how the additional model detail changes the investment results in the solution.

6 - Better Representative Period Selection in Capacity Expansion Optimizations Using Cheap Approximate Solutions

Nirmal Bhatt, Massachusetts Institute of Technology, Boston, MA, United States, Ruaridh Macdonald

Recent research has shown that capacity expansion optimizations give more robust solutions when they consider several years or decades of correlated weather and demand data. Similarly, stochastic optimization approaches allow extreme weather and events to be considered but also require optimizations to cover years of data. This need for long timeseries limits the feasible spatial resolution of models, meaning they do not capture important regional disparities, transmission congestion, and similar issues. One means of overcoming this challenge is to optimize energy systems over a subset of the timeseries, selecting representative periods which encompass the range of weather and demand patterns and lead to the same investment decisions and operational costs. This is a common method and can be combined with decomposition methods to parallelize the optimization and allow it to be distributed across a high-performance computer. So far, representative periods have been selected based on the model inputs but in this work we demonstrate that this can be improved by also incorporating approximate solutions to the investment and operations optimization. We show that our selection method allows for fewer or shorter representative periods to be used, reducing the total runtime even after calculating the approximate solutions and allowing for greater spatial resolution in the model, while also giving more accurate results.

MB76

Regency - 707

Power System Planning and Operation IV: Stochastic Energy Infrastructure Planning at Scale

Invited Session

ENRE: Other Energy

Chair: Rahman Khorramfar, MIT, Cambridge, United States

1 - Adversarial Generation of Uncertainty Realizations for Adaptive Robust Optimization

Aron Brenner, MIT, Cambridge, MA, United States, Rahman Khorramfar, Jennifer Sun, Saurabh Amin

Adaptive robust optimization is a powerful approach for planning under uncertainty that aims to balance costs of "here-and-now" first-stage decisions with those of "wait-and-see" recourse decisions made after uncertainty is realized. Chance-constrained variants of these problems aim to optimize first-stage decisions with consideration of a priori probabilistic guarantees on resulting second-stage costs and feasibility. In general, these models cannot be optimized exactly and must be solved using safe approximations, which can be extremely conservative in data-driven settings where the true distribution of uncertainty is unknown. In this work, we develop a column-and-constraint generation algorithm for a class of adaptive robust optimization problems that generates adversarial uncertainty realizations using a variational autoencoder (VAE). We provide uncertainty sets with corresponding probabilistic guarantees and perform worst-case analysis using a projected gradient approach over the VAE latent space that differentiates the optimal second-stage objective value with respect to uncertain parameters. Finally, we apply our approach to an adaptive robust capacity expansion model for a zonal representation of the New England power system and compare our results to existing methods for adaptive robust optimization.

2 - Impact of Improved Power Flow Representation on Stochastic Nodal Capacity Expansion Planning

Tomas Valencia Zuluaga, University of California at Berkeley, Berkeley, CA, United States, Amelia Musselman, Jean-Paul Watson, Shmuel Oren

More frequent and intense weather events have highlighted the importance of planning for a resilient power grid, underscored by increasingly weather-responsive electricity demand and generation. The challenge of handling these inherently uncertain risks can be addressed with stochastic nodal capacity expansion planning models that incorporate climate projection data. However, that comes at a high computational cost driven by the need for a finer geographic resolution. The use of power transfer distribution factors (PTDF) in the representation of power flow constraints can help in reducing the size of the resulting problem. It can also make it more suitable for scenario decomposition methods like the progressive hedging algorithm, but necessitates special consideration, since the topology of the system changes as part of the optimization process. In this talk, we extend an existing PTDF formulation for transmission expansion planning to our stochastic, joint generation, transmission and storage expansion planning model. We consider different linear model formulations with varying representation accuracy of nodal power flow constraints and analyze their impact on the resulting optimal cost and investment plan returned by the model, using a realistic 500-bus system as test case and a year of representative days.

3 - Stochastic Capacity Expansion Planning Considering Interconnection Queue Uncertainty

Elizabeth Glista, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, United States

Planning for the future of the power grid is of growing importance due to a variety of concerns: increasing demand for electric power, a shift toward weather-dependent sources of power generation, and increasing uncertainty due to climate risks. In this setting of high variability and uncertainty, multi-stage stochastic capacity expansion planning (CEP) can be used to inform long-term transmission, generation, and storage investment in a power system. However, its usefulness in practice is limited by the fact that most proposed expansion projects will never reach commercial operation. The overwhelming majority of proposed generation and storage projects are withdrawn at some point during or after the interconnection queue process. In this talk, we develop a probabilistic framework for quantifying the likelihood that a queued project will reach operational status, and we offer two distinct stochastic CEP variations that make use of these interconnection queue probabilities to inform investment decisions in the CEP model. By considering historical levels of queue clearing and construction, our models provide a more realistic description of possible future capacity expansion than existing CEP models. We compare results from the different nodal CEP model formulations on a large-scale synthetic-but-realistic California test case with over 8,000 buses and a year of representative future days. Our work demonstrates how the existing interconnection process limits CEP, providing further motivation for interconnection policy reforms.

4 - Integrated energy system planning under short-term and long-term uncertainty: Modelling and algorithms

Hongyu Zhang, Norwegian University of Science and Technology, Trondheim, Norway, Ignacio E. Grossmann, Brage Rugstad Knudsen, Ken McKinnon, Rodrigo Garcia Nava, Asgeir Tomasgard

We propose the REORIENT (REnewable resOuRce Investment for the ENergy Transition) model for energy system planning with the following novelties: (1) integrating capacity expansion, retrofit and abandonment planning, and (2) using multi-horizon stochastic mixed-integer linear programming with short-term and long-term uncertainty. We apply the model to the European energy system considering: (a) investment in new hydrogen infrastructures, (b) capacity expansion of the European power system, (c) retrofitting oil and gas infrastructures in the North Sea region for hydrogen production and distribution, and abandoning existing infrastructures, and (d) long-term and short-term uncertainty. We propose enhanced Benders decomposition methods to solve the model efficiently. We propose: (1) stabilising Adaptive Benders with the level set method and adaptively selecting the subproblems to solve per iteration for more accurate information, (2) a centred point stabilisation approach when the level method problem is hard to solve, and (3) dynamic level set management to improve the robustness of the algorithm by adjusting the level set per iteration.

The results show that: (1) compared with a traditional investment planning model, the REORIENT model yields 24% lower investment cost in the North Sea region, and (2) for a 1.00% convergence tolerance, the enhanced Benders is up to 6.8 times faster than the reference algorithm for MILP instances, and is up to 113.7 times faster than standard Benders and 2.14 times faster than unstabilised Adaptive Benders for LP instances. Finally, the dynamic level set management makes the algorithms more robust and is very helpful for solving large problems.

MB77

Regency - 708

Data-Driven Optimization for Reliable Machine Learning (I)

Invited Session

Computing Society

Chair: Vasileios Digalakis, HEC Paris, Jouy-en-Josas, France

Co-Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

1 - Distributionally Robust Optimization in Credible Learning with Healthcare Applications

Ioannis (Yannis) Paschalidis, Boston University, Boston, MA, United States, Yang Hu, Ruidi Chen

We consider the supervised learning of predictive models where predictor variables are organized into a group recommended by domain experts and several other groups of variables, where each other group contains variables encoding some specific type of information about the sample (e.g., indicator variables of some categorical predictor). We are interested in models that are dense in the expert group and group-sparse in the remaining groups, possibly allowing for the same variable appearing in multiple groups. We provide a distributional robust optimization formulation of this problem and a solution methodology. We demonstrate the effectiveness of our approach in several healthcare applications.

2 - Smart Predict-Then-Optimize for Two-Stage Linear Programs with Side Information

Alexander Estes, University of Maryland-College Park, College Park, MD, United States, Jean-Philippe Richard

We study two-stage linear programs with uncertainty in the right-hand side in which the uncertain parameters of the problem are correlated with a variable called the side information, which is observed before an action is made. We propose an approach in which a linear regression model is used to provide a point prediction for the uncertain parameters of the problem. We use an approach called *smart predict-then-optimize*. Rather than minimizing a typical loss function for regression, such as squared error, we approximately minimize the objective value of the resulting solutions to the optimization problem. We conduct computational tests that compare our method with other approaches for optimization problems with side information. The results indicate that our method can provide better objective values in situations where the true model is reasonably close to a linear model. Although the procedure we propose requires a longer time for fitting than existing methods, it requires less time to produce a decision for each given observation of the side information.

3 - An Extended Validity Domain for Constraint Learning

Yilin Zhu, University of Iowa, Iowa City, IA, United States, Samuel Burer

Many real-life optimization problems involve constraints that do not have explicit formulae, and learned constraints from available data will be used as an approximation. Therefore, we consider embedding a (predictive) machine learning model within a (prescriptive) optimization problem. In this setting, called constraint learning, we study the validity domain, a subset of the feasible set that constrains the optimization problem on a region where machine learning predictions are more reliable. In this talk, we propose a new validity domain which uses the convex hull idea in an extended space. We investigate its properties and compare it empirically with existing techniques on a set of test problems for which the ground truth of the optimization problems are known. Numerical results show that the extended convex hull method outperforms the existing methods in terms of function value errors. We also consider our approach within a pricing case study using real-world data.

4 - On a Rounding Mechanism for Orthogonality Constrained Optimization Problems

Ryan Cory-Wright, Imperial College London, London, United Kingdom, Jean Pauphilet

We propose a new rounding mechanism for orthogonality constrained problems, and study the quality of its solutions in both theory and practice. Orthogonality constraints arise in product recommendation applications like matrix completion, and data science applications line sparse principal component analysis with multiple components among others.

Joint work with Jean Pauphilet.

This is work in progress, stay tuned for a more detailed abstract nearer the conference.

MB78

Regency - 709

Decision Diagrams for Optimization

Invited Session

Computing Society

Chair: Anthony Karahalios, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Cutting Planes in Column Elimination

Anthony Karahalios, Carnegie Mellon University, Pittsburgh, PA, United States, Willem Jan van Hoeve

Column elimination is an exact method for solving discrete optimization problems. The method solved an open instance of the vehicle routing problem with time windows with 1,000 locations. To do this, column elimination solves a Lagrangian reformulation using a subgradient descent method. We study the incorporation of cutting planes in column elimination. We develop methodologies for adding both robust cuts and non-robust cuts from the column generation literature. For non-robust cuts, we implement weakened versions of the cuts and compare their strength to the weakened versions used in column generation. We compare different implementations and show experimental results on the capacities vehicle routing problem.

2 - Solving the Weapon Target Assignment Problem with Decision Diagrams

Ryan O'Neil, Nextmv, TBD, PA, United States

The Weapon Target Assignment Problem can be challenging to solve and to model. We examine this problem from the perspective of Decision Diagrams and Domain-Independent Dynamic Programming. We show this approach reduces modeling complexity and can yield good results.

3 - CODD: A Decision Diagram-based Solver for Combinatorial Optimization

Willem-Jan van Hoeve, Carnegie Mellon University, Pittsburgh, PA, United States, Laurent Michel

This presentation introduces CODD, a system for solving combinatorial optimization problems using decision diagram technology. Problems are represented as state-based dynamic programming models using the CODD language specification. The model specification is used to automatically compile relaxed and restricted decision diagrams that are embedded inside a branch-and-bound search process. We introduce abstractions that allow us to generically implement the solver components while maintaining overall execution efficiency. We demonstrate the functionality of CODD on a variety of combinatorial optimization problems and compare its performance to other state-based solvers as well as integer programming and constraint programming solvers. CODD provides competitive results and can outperform the other solvers, sometimes by orders of magnitude.

Monday, October 21, 12:45 PM - 2:00 PM

MC01

Summit - 320

E-commerce Supply Chain Operations

Invited Session

Service Science

Chair: Ting Wang, University of Science and Technology of China, Hefei, N/A

1 - From Storefronts to Screens: An Exploration of Ship-From-Store, Customer Behavior, and Sku-Level Impacts

Hsiao-Hui Lee, National Chengchi University, Taipei, Taiwan, Wenzheng Mao, Zhanyu Dong, xiaoli Liu

In the ongoing evolution from brick-and-mortar (B&M) stores to multichannel operations, the “Ship-from-Store” (SFS) strategy stands out as a pivotal approach for achieving seamless omnichannel integration. Through SFS, B&M stores harness their in-store inventory to swiftly cater to the online demands of proximate customers. However, despite its potential to broaden customer reach, inappropriate implementation of SFS may yield detrimental effects for retailers, such as diminished cross-selling opportunities. Leveraging empirical data from a pharmacy retailing chain, this paper embarks on an investigation into the effects of SFS on customer behavior, inventory management, and resultant profitability. The paper's data-driven insights particularly underscore the significance of SKU-level analysis in comprehending the multifaceted implications of SFS on retail operations. With empirical observations as a foundation, we construct a stylized model and find the conditions under which a B&M retailer can benefit from SFS-namely, when in-store visit costs outweigh online waiting costs, thereby promoting a demand pooling effect and enhancing in-store fill rates. The study further reveals that reduced online waiting costs through SFS can induce market expansion or channel cannibalization, with the former proving the most profitable trajectory. In scenarios of partial channel cannibalization, profits can be augmented by astutely selecting high-margin, high cross-selling, and well-stocked items for online listing. Finally, our empirical analyses substantiate the paper's theoretical postulations and results, providing a cohesive narrative on the dynamics of SFS in modern retail.

2 - How Retail System Complexity Drives Inventory Record Inaccuracy: Empirical Evidence from Cross-Border E-Commerce

Ting Wang, University of Science and Technology of China, Hefei, China, People's Republic of, Kejia Hu, Stanley Lim, Yun Fong Lim, Yugang Yu

Problem definition: Accuracy of inventory recording stands as a fundamental pillar for maintaining and enhancing operational efficiency within retail systems. However, the advent of e-commerce presents an escalated challenge to this accuracy by exposing e-retailers to complex retail systems. Despite previous focus on internal complexity in offline retailing, understanding of how external complexity influences inventory record inaccuracy (IRI) remains scarce. This paper broadens the scope from traditional retail settings to e-commerce, offering an empirical exploration of the diverse dimensions of retail system complexity and their impact on IRI.

Methodology/Results: Utilizing a proprietary panel dataset from a cross-border e-retailer, inclusive of inventory, sales, and product attribute data, we delineate retail system complexity into various dimensions (internal versus external, static versus dynamic, and product versus component). We employ random-effects models to probe the relative significance of these dimensions in predicting IRI and use an instrumental variable approach to explore their causal effects on IRI. Our results underscore the overlooked role of external complexity as a more potent predictor of IRI in e-commerce, both independently and through its amplification of internal complexity's effects on IRI.

Managerial implications: Our findings lead to actionable insights into how e-retailers can strategically manage their retail system complexity to mitigate IRI and enhance profits. While system complexity can erode profit through IRI, it can also augment profit by attracting greater demand. Our counterfactual analysis yields specific guidelines for e-retailers to leverage the dual nature of retail system complexity for profit maximization.

3 - Supply Chain Instant Retail Introduction Strategy Considering Consumer Time Value

Sichao Yang, Tianjin University, Tianjin, China, People's Republic of, Baofeng Huo, Lin Miao

Online retailers and brick-and-mortar stores have collaboratively developed an "order online, fulfill in-store" instant retail service, leveraging online traffic and physical store capabilities to meet consumers' immediate needs, thereby creating a novel retail model. This study examines the introduction strategies of instant retail channels based on online channels of e-retailers and offline channels of physical retailers. It incorporates the quantified time value of consumers into channel selection, analyzing how time value, delivery timeliness, and commissions impact the competitive and cooperative relationships and game-theoretic equilibrium between channels. The findings indicate that the introduction of instant channels significantly affects the pricing competitiveness of both online and physical retail channels. When consumers' sensitivity to time value is low, the instant channel exhibits a clear competitive advantage, leading to a decline in market shares for both online and offline channels. Conversely, when consumers' time value sensitivity is high, the competitiveness of the instant channel diminishes, thereby enhancing the market shares of online and offline channels. Additionally, as consumers' sensitivity to time increases, the pricing and profits of instant and offline channels improve, though this change may adversely affect the market demand for the instant channel. Meanwhile, the additional costs of offline channels and consumers' perceptual differences in channel value significantly influence operational strategies. Overall, by implementing a reasonable instant channel strategy, it is possible to optimize channel efficiency and enhance the profitability of the entire supply chain without undermining the interests of offline channels, thereby achieving a Pareto improvement in the supply chain system.

4 - Exploring the Impact of Data Sharing and Direct Sales Channels on Corporate Decision-Making: A Perspective on Data Value

chunshen wang, Tianjin University, Tianjin, China, People's Republic of, Sichao Yang, Lin Miao, Baofeng Huo

With the rapid evolution of the digital economy, data resources have significantly become a key asset for companies to gain competitive advantage. In various links of the supply chain, the sharing of data has produced important new influences on product sales and service decisions. To deeply explore the interaction between enterprises' data sharing decisions and channel selection strategies, we constructed a game theory model involving a single manufacturer and a single retailer. The research findings indicate that in traditional retail channels, the decision of the manufacturer and retailer to share data is primarily influenced by both the cost of data sharing and the value of the data itself. In scenarios without direct sales channels, lower sharing costs prompt both parties to lean towards sharing data, while higher costs lead to a choice to retain data. After introducing direct sales channels, the decision-making process becomes more complex, especially in the context of higher sharing costs, leading to divergent strategic choices between manufacturers and retailers. Moreover, this study also finds that, whether under data sharing or retention strategies, manufacturers consider introducing direct sales channels only when the data has high mining value, whereas retailers tend to prefer maintaining traditional single channels. This research provides profound management insights into data sharing and direct sales channel strategies, emphasizing the importance of the unit value of data in decision-making, and the significant impact of data sharing strategies on enterprise profitability in a data-driven business environment.

MC02

Summit - 321

QSR Student Introduction and Interaction

Panel Session

Quality, Statistics and Reliability

Co-Chair: Akash Deep, Oklahoma State University, Stillwater, United States

1 - Moderator Panelist

Akash Deep, Oklahoma State University, Stillwater, OK, United States

2 - Panelist

Yu Jin, Unversity at Buffalo, Buffalo, NY, United States

3 - Panelist

Panelist

4 - Panelist

Panelist

5 - Panelist

Panelist

6 - Panelist

Panelist

MC03

Summit - 322

Recent Advances in Design and Analysis of Computer Experiments

Invited Session

Quality, Statistics and Reliability

Chair: Cheoljoon Jeong, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Upala Junaida Islam, Arizona State University, Tempe, AZ, United States

1 - Selective Parameter Calibration via Sliced Sequential Design

Cheoljoon Jeong, University of Michigan, Ann Arbor, MI, United States

Parameter calibration seeks to estimate unobservable parameters in a computer model by aligning field observations with computer model outputs. In the building sector, a physics-based building energy computer model is developed to analyze the building energy use, given various weather conditions and operational scenarios. To obtain accurate simulations, it is necessary to calibrate building-related parameters required for pre-configuration. Recently, Bayesian optimization has received attention in the literature, but it suffers when the number of parameters is large since it tends to overly search the parameter space near the boundaries where predictive uncertainty is high. It becomes even more problematic when model misspecification occurs in these boundary regions, yielding unreasonable outcomes. To address this challenge, this study suggests a new calibration approach that slices the parameter space stochastically based on the importance measure and that designs the next query point in the respective subspaces sequentially. The numerical studies and building energy simulation case study demonstrate that the proposed approach achieves a significant improvement in both calibration accuracy and efficiency, compared to other alternatives. Furthermore, our approach allows to explain the parameter importance, so that it examines which parameters receive greater emphasis throughout the calibration process.

2 - Complete Active Learning for Emulation and Optimization

Difan Song, ISyE Georgia Tech, Atlanta, GA, United States, Roshan Joseph

Gaussian process (GP) models are widely used in active learning for emulation and optimization of black-box functions. The existing GP-based active learning procedures start with an initial design and then add points using some acquisition functions. If the initial design is too small, the response surface may be under explored and the algorithm may terminate prematurely at a local optima. On the other hand, if the initial design is too large, then we may waste valuable resources and may miss the interesting regions of the response surface. This article proposes a new active learning procedure that completely avoids the initial design. This is achieved by using a new correlation function and a new GP model, which automatically embeds a projection-based space-filling criterion into the acquisition functions. Through theory and simulations, we show that the proposed procedure, which we call COMPlete ACTive (COMPACT) learning, outperforms the existing active learning procedures.

3 - Consistent Exploration-Exploitation Trade-Off in Active Learning Regression with Dirichlet Process Prior

Ashif Iquebal, Arizona State University, Tempe, AZ, United States, Upala Junaida Islam

The trade-off between exploration and exploitation in active learning for regression problems remains significantly understudied. This paper addresses this challenge by generalizing a Bayesian hierarchical model to estimate the trade-off parameter η dynamically. We model η using a mixture distribution with a Dirichlet process prior, effectively capturing uncertainty within and between querying stages. Our work is the first to provide theoretical guarantees on the convergence and consistency of the exploration-exploitation trade-off, as well as bounds on the sample complexity of the active learning framework. We validate our approach through simulated and real-world case studies, demonstrating an average 14% reduction in generalization error compared to traditional active learning methods.

4 - Bayesian Optimization for Knowledge Discovery in Electromagnetics

Pouya Ahadi, Georgia Institute of Technology, ATLANTA, GA, United States, Kamran Paynabar, Reza Marzban, Ali Adibi

Electromagnetic (EM) nanostructures, such as metamaterials and metasurfaces, play a crucial role in advancing the compactness and functionality of modern devices by controlling light-matter interactions at subwavelength scales. Understanding and optimizing these interactions is key to fully utilizing their potential, yet traditional design methods are often hindered by extensive computational demands and limitations in accurately capturing the complex dynamics involved. The primary goal is to enhance device performance through novel designs that effectively control electromagnetic wave propagation, which is crucial for developing next-generation technologies.
Artificial intelligence (AI)-based methods have been used to enhance design optimization, providing powerful tools for predicting and optimizing the behavior of nanostructures. Despite their advantages, these techniques often face challenges with non-unique design scenarios and can be complex to implement and understand. In this study, we use the Gaussian Process (GP) regression technique to build a relationship between the design of a nanostructure and its wave propagation, as measured by its reflection spectrum. Considering the time-consuming nature of experimental measurements of wave propagation for a specific design, we face the challenge of limited data availability. We apply Bayesian Optimization (BO) to explore near-optimal design patterns, aiming to provide practical improvements in the field of electromagnetic nanostructures. Our experimental results validate the effectiveness of our proposed BO framework in achieving improved design solutions.

5 - Accelerating Material Discovery through Balanced Exploration-Exploitation of the Design Space

Ahmed Shoyeb Raihan, West Virginia University, Morgantown, WV, United States, Imtiaz Ahmed

The quest for discovering new materials with optimal properties is a critical endeavor in advancing technological and environmental progress. Our research addresses the critical challenge of balancing exploration and exploitation within the domain of accelerated material discovery, employing sequential learning techniques. The exploration-exploitation dilemma is a major challenge in material discovery and development, where exploration refers to the search across a broad range of potential material configurations, often leading to novel discoveries, and exploitation involves focusing on refining known materials to optimize their properties. Efficiently managing this trade-off is crucial as excessive exploration can deplete resources without guaranteeing results, whereas excessive exploitation might overlook potentially superior materials. Our approach is designed to navigate the extensive material design space efficiently, aiming to quickly hone in on materials with the desired properties while minimizing experimental resources. We propose a strategic framework that improves upon traditional methods by better managing the exploration-exploitation balance, thereby significantly reducing the time and cost associated with experimental trials. This study is expected to offer insights into more effective strategies for material discovery, enhancing the speed and efficiency of identifying materials with desired properties across various domains.

Keywords: Sequential Learning, Accelerated Material Discovery, Exploration-Exploitation, Bayesian Optimization

MC04

Summit - 323

High-dimensional Data Analysis: Modeling, Monitoring and Control

Invited Session

Quality, Statistics and Reliability

Chair: Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

1 - Robust multivariate functional tensor decomposition

Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

In the landscape of Industry 4.0, a vast array of data is captured during manufacturing processes, frequently contaminated with outliers that can severely weaken the performance of process monitoring procedures, particularly in complex and high-dimensional environments. Addressing this, we propose a robust functional tensor decomposition algorithm to enhance monitoring processes by tackling time decentralization and anomaly contamination effectively. Signal data, whether from manufacturing sensors or health monitoring devices, commonly encounters the dual challenges of decentralized time recording and universal anomalies across time and feature domains. These complications can significantly undermine the efficacy of traditional control charting systems. Our novel algorithm strategically decomposes the observed functional tensor into two distinct components: the true low rank functional tensor and an anomalous tensor. This decomposition is achieved by leveraging the continuity characteristic in functional signals and exploiting the sparsity of anomalies, thus ensuring a more accurate anomaly detection and isolation. To validate our approach, we employ a simulation framework that demonstrates the algorithm's enhanced capability to identify and separate anomalies within sparse tensors. This capability marks a significant improvement over conventional methods that typically neglect the functional nature of data or struggle with cell-wise or component-wise anomalies. The results not only emphasize a theoretical advancement in functional data analysis but also highlight practical implications for industries reliant on precise and reliable data monitoring systems.

2 - Optimizing Machine-to-Machine Reproducibility with Deep Auto-Encoding Multi-Task Gaussian Processes

Logan Heck, The University of Texas at San Antonio (UTSA), San Antonio, TX, United States, Adel Alaeddini, Krystel Castillo

Manufacturing industries continually seek innovative approaches to enhance product quality in a cost-effective manner. Machine learning algorithms are proficient in recognizing complex features and patterns in large swaths of data. This study introduces a deep auto-encoding multi-task Gaussian process to improve manufacturing reproducibility for parts produced using a Form 3+ SLA 3D printer. Building on prior research that captured input parameters and dimensional deviations for the selected SLA printer, this study leverages the proposed Gaussian process to predict the resulting deviations on parts based on specific input parameters. This utilization of a machine learning algorithm for optimization purposes will allow for a less time-consuming method in determining what input parameters significantly influence the overall quality of parts produced via additive manufacturing. While limited by a small number of available data points, this study showed promising results in the algorithm's ability to better predict the true deviations of the printed parts over time.

3 - Sparse Linear Discriminant Analysis with Knowledge Transfer

Xinwei Deng, Department of Statistics, Virginia Tech, Blacksburg, VA, United States, Xiaoning Kang, Xiaoyu Chen

High-dimensional data with limited sample size commonly arise in the scientific and engineering applications. The data scarcity brings much difficulty for training models and conducting inference. The transfer learning is known to be a useful technique when different but possibly related data (auxiliary data) are available in addition to the original target data in the study for modeling and estimation. In this work, we propose a transfer learning method for sparse linear discriminant analysis under the high-dimensional data setting. The key idea of the proposed method is to effectively identify the informative auxiliary data and transfer the useful information from such selected auxiliary data to greatly enhance the modeling of the target data. Theoretical properties of the proposed method are investigated coupled with developing an efficient computational algorithm for model estimation. The performance of the proposed method is evaluated by both simulation and real case study.

4 - Predictive modeling of surface topography using power signal in wire electrical discharge machining

Bochi Liu, Rutgers, The State University of New Jersey, Piscataway, NJ, United States, Yuebin Guo, Weihong "Grace" Guo, Wenyuan Wang

Wire electrical discharge machining (WEDM) is a non-contact manufacturing process that uses a travelling wire electrode to cut hard workpieces with complex geometries. Surface topography significantly impacts the performance of machined workpieces, affecting qualities such as fatigue life, abrasive wear, and corrosion resistance. Measuring the surface topography of machined workpieces usually needs to be done after the machining process is complete, and some measurements may cause irreversible damage to workpieces. Therefore, online reconstruction of surface topography in WEDM is necessary to improve surface quality and production efficiency. Furthermore, as WEDM gets smarter and more complex, it involves a large number of control parameters, such as peak current, voltage, pulse on time, pulse off time, and electrode feed rate, etc. Current physics-based methods require complex mathematical modeling of the WEDM process, which is often challenging to implement. Machine Learning (ML) methods can estimate the complicated relationship between control parameters and surface topography based on data, which reflects their advantages compared with physics-based methods. Additionally, since WEDM relies on electrical discharges to remove material, it is promising to simplify the high-dimensional control parameters using WEDM's discharge power. This study proposes a predictive modeling of surface topography based on power signal in WEDM. Time-frequency analysis and empirical mode decomposition are used to extract features from the power signal and measured surface topography. Multiple ML models are employed to find the complex nonlinear relationship between these features. Spatial interpolation analysis is applied to reconstruct the predicted surface topography.

MC05

Summit - 324

Diversity, Equity, and Inclusion Research in Operations Management

Invited Session

Diversity, Equity, and Inclusion

Chair: Lu Kong, University of South Florida, Tampa, FL, United States

1 - Bias Beyond the Screen: An Inequity Wake-up Call in Online Teaching's Evaluations

Lu Kong, University of South Florida, Tampa, FL, United States, Kejia Hu, Lanfei Shi, Timothy Vogus

Our study investigates the impact of the transition to online education on student evaluations of teaching (SETs), with a particular focus on disparities related to the gender and race of instructors. Utilizing data from RateMyProfessor.com, which includes 598,252 ratings from instructors at the top 150 U.S. colleges, we employ a Difference-in-Differences-in-Differences (DDD) model design combined with Propensity Score Weighting to analyze these effects. This approach helps us quantify changes in SETs following the shift from offline to online teaching environments, particularly during the first six months after U.S. universities adopted online formats due to COVID-19.

Our findings reveal a pronounced decline in evaluations for non-white male instructors, who experienced a 27.03% drop in ratings—a significantly steeper decrease than their counterparts. Further analysis using text mining techniques on student comments suggests that this disparity is partly driven by reviews containing language sensitive to diversity and inclusion, as well as comments emphasizing the hedonic aspects of education.

The study highlights the underlying mechanisms contributing to biases in SETs and suggests that these disparities are exacerbated in online learning contexts. To address these issues, we propose several strategies for educational institutions. These include the adoption of multiple assessment methodologies to evaluate teaching effectiveness, the redesign of teaching evaluation questionnaires to eliminate inherent biases, and initiatives to raise awareness about potential biases among students and faculty.

Our research provides valuable insights into the challenges of fair evaluation in online education and offers practical recommendations for enhancing equity in academic assessments.

2 - Racial Disparities in Health Technology and Its Impacts on Quality of Care: the Case of Biased Readings from Spo2 Sensors

Gaurav Jetley, Colorado State University, Fort Collins, CO, United States, He Zhang

Recent explorations in healthcare processes have revealed a significant bias in clinical measurements affecting care delivery, particularly within Intensive Care Units (ICU). Notably, pulse oximetry has been identified as systematically overestimating oxygenation levels in Black patients compared to White patients. This study investigates how such biases in SpO2 measurements influence subsequent clinical and operational decisions, thereby impacting patient outcomes and resource utilization in ICUs. Utilizing a comprehensive dataset of approximately 73,000 ICU stays, patients were divided into those monitored with only SpO2 and those with both SpO2 and SaO2 measurements. We use a serial mediation model which focuses on the impact on early readmissions due to early discharges taking place because of lower supplemental oxygen delivery likelihood due to inflated SpO2 measurements for Black patients. Our causal identification strategy using instrumental variables shows that Black patients in the SpO2-only group were less likely to receive supplemental oxygen, which leads to shorter LOS and higher 30-day readmission rates. This relationship was mediated by the omission of supplemental oxygen and a shortened LOS, suggesting a risk of premature discharge. Conversely, the availability of both SpO2 and SaO2 measurements neutralized racial disparities in oxygen delivery and early discharge. The findings underscore the need for refined healthcare protocols and information systems to identify and protect high-risk patients, ensuring equitable healthcare delivery across all demographic groups. These operational insights are crucial for improving quality of care and reducing readmissions in racially diverse populations.

3 - Can Workplace DEIB enhance manufacturing environmental performance? Evidence from Employment Non-Discrimination Acts

Suvrat Dhanorkar, Pennsylvania State University, University Park, PA, United States, Max Ji, Suresh Muthulingam, Kevin Linderman

In this study, we explore how enhanced workplace Diversity, Equity, Inclusion, and Belonging (DEIB) affects toxic releases in manufacturing operations. To investigate this idea, we utilize the staggered adoption of Employment Non-Discrimination Acts (ENDA), which prohibit discrimination based on gender identity or sexual orientation. We adopt a robust staggered difference-in-differences (SDiD) approach, which allows for heterogeneous effects and a conditional parallel trend assumption, and find that enhanced workplace DEIB, facilitated by ENDA enactments, leads to better environmental performance.

4 - Racial and Gender Biases in Customer Satisfaction Surveys: Evidence from a Restaurant Chain

Yong-Pin Zhou, University of Washington, Seattle, Seattle, WA, United States, Masoud Kamalahmadi, Qiuping Yu

Racial and gender inequalities are ubiquitous in the workplace. Whereas previous studies have primarily focused on employer discrimination, we study the role of customers, using 1,444,044 transactions and 257,656 customer satisfaction surveys from a full-service casual-dining restaurant chain in the US. We find that customer ratings of servers are biased against racial minority, and, interestingly, also against females despite their majority in this occupation. We further find evidences that suggest that statistical discrimination is the primary driver for racial biases, while status-based discrimination is likely the main driver for gender biases. These findings are robust in a number of additional tests. Given these underlying mechanisms, we propose tailored strategies to mitigate the biases.

MC06

Summit - 325

Pricing and Selling Strategies

Contributed Session

Chair: Mitsuki Nakano, Waseda University, Tokyo, N/A, Japan

1 - Foot-in-the-door auctions: Surcharges in online procurement

Timo Heinrich, Hamburg University of Technology, Hamburg, Germany, Ernan Haruvy, Matthew Walker

Anecdotal evidence suggests that buyers in online procurement are at the risk of hold-ups despite reputation mechanisms and possible litigation. In a first step, we analyze data from a popular platform that offers buyer-determined procurement auctions for services. Based on a structural econometric model, we provide evidence that some bidders ask for very low prices that can be rationalized through unexpected surcharges. We assume that negative costs are recovered through unobserved surcharges and estimate a lower bound for their effect on buyer surplus. We find that for medium-cost and high-cost projects, the effect is negligible. However, for low-cost projects, surcharges amount to at least a quarter of the final price, which reduces expected buyer surplus. In a second step, we conduct a pre-registered online experiment to evaluate which aspects of the decision environment drive surcharging. Many factors can affect surcharging in the field, including widely studied aspects such as the cost overruns, expected cost of litigation or reputational concerns. We develop a stylized procurement game and disentangle the role of three additional factors: fairness norms, social norms, and competition.

2 - Interactions between the platform's entrant and manufacturer's channel decisions under the dual-market system

Ruiying Yuan, Tianjin University, Tianjin, China, People's Republic of, Zhaofang Mao, Zuojun Max Shen

The full lifecycle concept has prompted sellers to provide ancillary services based on traditional product sales, leading to a dual-market system consisting of a base market and an add-on market. In this study, we consider a manufacturer selling base products through an online retail platform and then selling add-on products directly to consumers who have purchased base products. We investigate how the manufacturer's distribution channel strategy in the base market interacts with the platform's entrant strategy in the add-on market. We identify the manufacturer's wholesale price effect (sales-control effect) in the base market caused by the platform's entrant of the add-on market under the reselling (agency) channel. Additionally, when facing negative influences, the dual-market system endows the manufacturer with the focus-shift effect. If the manufacturer adopts the agency (reselling) channel in the base market, the platform prefers (not) to enter the add-on market to compete with the manufacturer; if the manufacturer adopts the dual-channel, the platform prefers to enter only if both the commission rate and channel competition are high. Furthermore, the manufacturer prefers the dual-channel when both the commission rate and channel competition are low. Interestingly, due to the interactions between the two firms, the manufacturer will adopt the agency channel instead when the commission rate is extremely high. Finally, we examine conditions under which the platform has incentives to allow the manufacturer to change from a single-channel to a dual-channel in the base market.

3 - Dynamic Pricing Algorithm with Demand Learning Under Minimax Regret

Jingcheng Lei, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, ruijing wu, Shaoxuan Liu

We consider a monopoly firm that sets prices for a new product dynamically under incomplete information. At the beginning of the selling season, the firm only knows the support of customers' valuation, but has no information about the distribution of their valuation. In order to model this situation, we design a robust dynamic pricing scheme whose objective is to minimize the worst-case regret. By observing customers' purchase actions, the firm learns customers' valuation continuously and updates their pricing decisions dynamically. We consider two distinct kinds of customers: homogeneous customers and heterogeneous customers. For homogeneous customers, we design an optimal pricing policy with respect to the minimax total period regret rule for each possible scenario. We show that the minimax regret is identical in any price paths and prove that the T-period minimax regret is bounded by O(log T). For heterogeneous customers, given information on historical prices, we propose an forward algorithm of calculating optimal price under minimax regret. Furthermore, numerical experiments show that the algorithm can accurately learn the optimal price and performs better compared to the traditional methods.

4 - Pricing Non-Standard Vegetables Considering Customer Evaluation

Mitsuki Nakano, Waseda University, Tokyo, Japan, Kotomichi Matsuno, Takahiro Ohno, Yoshikuni Edagawa, Takaaki Kawanaka, Shinya Takata

In Japan, approximately 1.7 million tons of harvested vegetables are not shipped, accounting for about 13.7% of the total. Additionally, some vegetables are discarded before being counted in the harvest. Including these discarded vegetables, it is reported that approximately 30% of the vegetables produced are wasted. These vegetables that do not meet shipping standards are called "non-standard vegetables." Non-standard vegetables are produced by farmers that do not meet the standards set by agricultural cooperatives due to damage, size, shape, etc., and cannot be sold on the market.

Currently, an increasing number of farmers are selling through D2C channels such as e-commerce sites to reduce the loss of non-standard vegetables. However, interviews with e-commerce businesses that sell substandard vegetables revealed that demand is unstable. This may be

due to incorrect pricing.

The purpose of this research is to propose a method for calculating the price that consumers feel is appropriate for non-standard vegetables when compared with standard vegetables. A prediction model is constructed using the appearance characteristics (blemishes, size, shape, color) and consumer attributes of non-standard vegetables as explanatory variables, and the expected purchase price as the objective variable. Based on the vegetable standards set by agricultural cooperatives, the appearance characteristics of non-standard vegetables are divided into two to three levels. An experimental design method is used to conduct an online survey asking consumers what a fair price is. The results of this research are expected to lead to methods that increase consumers' willingness to purchase non-standard vegetables at reasonable prices.

MC07

Summit - 327

Facility Management and Network Design

Flash Session

Chair: ZHENGHAO SUN, N/A

1 - A Study on the Successful Opportunity of Regional Revitalization in Old Public Dormitories: A Dynamic Perspective from Actor-Network Theory

wenjung Chang, National Chung Hsing University, Taichung, Taiwan, Chang-Yen Lee, Mingchang Chih, Yu-Jyun Ma

In recent years, with the rapid development of urbanization, countries around the world are facing the problem of unbalanced urban and rural development, with the decrease and aging of rural populations becoming increasingly prominent. Therefore, areas outside of cities need to actively seek pathways for revitalization, cultural preservation, and sustainable development in the future. Governments worldwide are paying more attention to balanced urban and rural development, with local innovation becoming a popular keyword for regional revitalization.

The article adopts an Actor-Network Theory (ANT) perspective to explore the dynamic evolution of the life cycle of the " regional revitalization" industry across four stages: issue formation, organizational networks, productivity, and redevelopment, as well as decline and dissipation. Through this exploration, it aims to observe the dynamic developmental trends of network governance among organizations.

Through an open-convergent exploratory research approach, the study aims to identify the issues in network governance at each stage within an integrated framework, along with the key influencing factors on implementation. Additionally, employing independent sample t-tests and Cohen’s d to test effect sizes, the study conducts independent sample heterogeneity difference analysis between the PR and BM groups, revealing significant cognitive differences regarding the success opportunities for the revitalization of Zhongxing New Village street blocks at the "Organizational network stage". The findings contribute to broadening the theoretical perspective on inter-organizational network governance relationships, while also proposing research conclusions and practical strategy recommendations for the development of organizational networks in revitalizing old military dependents’ village street blocks.

2 - Assessment and Industry Benchmarking of Supply Chain Resilience to Understanding Resiliency Capabilities in Specific Industries in Singapore

Ning Liu, Advanced Remanufacturing & Technology Centre (ARTC), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore, Allan N Zhang, Louiz Lee, Shanshan Yang

Within the dynamic landscape of Singapore, the imperative of Supply Chain Resilience (SCR) amidst global COVID-19 pandemics, natural disasters, and geopolitical tensions, is critical for businesses. This paper presents a combination of qualitative and quantitative methodologies for resilience assessment and benchmarking in specific industries. SCR, the capacity to withstand disruptions while consistently delivering goods and services efficiently, is investigated through seven crucial capabilities, including agility, anticipation, collaboration, diversification, flexibility, redundancy, and visibility. To measure SCR, a resilience score (R+) ranging from 1 to 5 is proposed for benchmarking.

This paper presents some key findings as follows. Firstly, it offers insights into the companies’ SCR status at enterprise level, identifying gaps and providing a roadmap for bolstering SCR. Secondly, we investigate the industry landscape of Singapore, highlighting differentials and gaps in SCR capabilities across sectors. Notable resilience variations across industries and significant differences between the top and bottom tiers are also explored. Thirdly, by illuminating industry-specific resilience dynamics, we contribute to the scholarly understanding of SCR in the Singaporean context by benchmarking against peer performance. Finally, this paper provides a systematic approach for assessing SCR including identifying gaps, prioritizing necessary capabilities, and creating actionable plans for practitioners and policymakers. Tailored approach designed to address the unique challenges encountered by each sector are recommended to foster the adoption of robust SCR practices throughout Singapore's supply chains.
This research is supported by the Agency for Science, Technology and Research (A*STAR) under its A*STAR – MiRXES Agile Diagnostics Manufacturing Joint Lab (Grant No.: I2301E0017)

3 - A unified modeling framework of mixed traffic based on macro-micro behavioral connection

Jia Li, Washington State University, Pullman, WA, United States, Di Chen

Traffic systems are shared by heterogeneous driving agents (e.g. human drivers and autonomous vehicles, passenger cars and trucks), therefore modeling mixed traffic plays a pivotal role in understanding, managing, and operating such systems. Various mixed traffic models have been proposed in recent years, but their context and rationales vary considerably from each other. There lacks a unified framework to connect and coherently explain them, and this limited understanding hinders the interpretability and generalization of these models. In this presentation, we propose a unified analytical modeling framework to fill this gap and show how the framework helps to better understand

mixed traffic behaviors. The core idea of this framework is establishing the connection between macroscopic flux functions of mixed traffic with microscopic agent interactions, encapsulated by their speed functions and cooperativeness. We show that interactions of agents will fully determine a mapping between microscopic perceived density and macroscopic density of the system. This property allows a unified derivation of macroscopic flux functions of mixed traffic, which reduces to computing microscopic perceived densities and estimating speeds of different classes of agents. Using this framework, we will discuss the connections of a few popular models of mixed traffic and uncover microscopic behaviors of driving agents implied by these models and highlight the role of driving agent interactions in achieving collective cooperation in mixed traffic systems.

4 - Commuters' periodic mode usage pattern in the credit-based mobility system with trip uncertainty

HONGXING DING, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Xinwei Li, Hai Yang

Uncertainty has received great attention in traffic demand management, which is commonly considered from an aggregate viewpoint. Few studies investigate tip uncertainty that arises from individual future uncertain trip needs. This study examines heterogeneous commuters’ day-to-day strategic mode usage behavior during a credit Charge-cum-Reward (CCR) scheme period, considering individual trip uncertainty. We consider whether commuters are confronted with an irregular and urgent trip as a random variable which is known only to the individual at the start of each day. Based on such trip type information, each commuter decides the mode usage behavior for both current day and future mode usage pattern to minimize individual periodic travel cost. We model commuters’ day-to-day mode choices as a two-stage mixed-integer linear programming problem. The model provides a closed-form solution for each commuter’s daily best strategy based on the current trip type and credit balance. The day-to-day variances in both individual mode usage behavior and system-wide flow pattern across a CCR scheme period are further examined. We also explore the system properties and stability conditions. Our findings show that the system-wide auto flow exhibits two trends across the period, with relatively high flow at the beginning stage and relatively low flow at the second stage. The trend change point is related to the CCR scheme. We reveal the impact of the period length on commuters’ daily decisions and provide insights into how individual commuters make optimal day-to-day dynamic mode choices under the CCR scheme with trip uncertainty.

5 - Optimal Allocation of Critical Equipment in Collocation Data Centers

Neda Zarayeneh, Hitachi America, Santa Clara, CA, United States, Malarvizhi Sankaranarayanasamy, Pegah Mavaie , Omanshu Thapliyal, Prasun Singh, Ravigopal Vennelakanti

In the digital era, colocation data centers support businesses of varying sizes but need help with the supply backlog of critical equipment, delaying the fulfillment of customer demands. To address this issue, this paper presents a comprehensive approach encompassing all stakeholders within the data center's supply chain, including demand, supply, procurement, and operational processes. We emphasize the importance of forecasting demands well in advance to enable timely procurement of essential equipment. We integrate internal and external demand signals to develop precise, long-term, multi-horizon forecasting models.

Our study introduces a novel multi-variable time series forecasting framework designed to predict the power capacity needs of colocation data center customers. This framework uses a modified transformer-based model, PatchTST, customized for power capacity predictions. We evaluate its performance against RNN-based models like SegRNN and LSTM and transformer-based models like FEDformer and Informer, demonstrating its effectiveness in long-term demand forecasting.

Current industry practice of predicting demand based on the past two to three quarters often leads to inflated or inaccurate forecasts influenced by sales personnel, a phenomenon known as sandbagging. Our method incorporates external factors like macroeconomic indicators to enhance forecasting precision. After establishing demand, we align predicted capacities with required equipment and manage procurement. We then optimize resource allocation to meet customer demands, minimize wait times, and enhance the order-to-cash cycle. A dynamic system tracks remaining customer capacity and prioritizes resources effectively. The problem is mathematically formulated to reduce delivery times and increase revenue, constrained by facility capacity, demand sufficiency, resource availability, and non-negativity.

MC08

Summit - 328

Working in Military and Security

Panel Session

Military and Security

Co-Chair: Isabella Sanders, Dept. of Systems Engineering, United States Military Academy, Newburgh, United States

1 - Panelist

Jesse Pietz, MIT Lincoln Laboratory, Lexington, MA, United States

2 - Panelist

Isabella Sanders, United States Military Academy at West Point, West Point, NY, United States

3 - Panelist

Nathaniel Bastian, United States Military Academy, Chester, NY, United States

4 - Panelist

Carson Long, Air Force Institute of Technology (AFIT), Dayton, OH, United States

5 - Panelist

Adam Haywood, United States Air Force, Pentagon, VA, United States

MC09

Summit - 329

Learning in Online Platforms

Invited Session

Auctions and Market Design

Chair: Jackie Baek, New York University, New York City, NY, United States

1 - Improved Bayes Risk Can Yield Reduced Social Welfare Under Competition

Meena Jagadeesan, UC Berkeley, Berkeley, CA, United States, Michael Jordan, Jacob Steinhardt, Nika Haghtalab

As the scale of machine learning models increases, trends such as scaling laws anticipate consistent downstream improvements in predictive accuracy. However, these trends take the perspective of a single model-provider in isolation, while in reality providers often compete with each other for users. In this work, we demonstrate that competition can fundamentally alter the behavior of these scaling trends, even causing overall predictive accuracy across users to be non-monotonic or decreasing with scale. We define a model of competition for classification tasks, and use data representations as a lens for studying the impact of increases in scale. We find many settings where improving data representation quality (as measured by Bayes risk) decreases the overall predictive accuracy across users (i.e., social welfare) for a marketplace of competing model-providers. Our examples range from closed-form formulas in simple settings to simulations with pretrained representations on CIFAR-10. At a conceptual level, our work suggests that favorable scaling trends for individual model-providers need not translate to downstream improvements in social welfare in marketplaces with multiple model providers.

2 - Social Learning with Bounded Rationality: Negative Reviews Persist Under Newest First

Atanas Dinev, Massachusetts Institute of Technology, Cambridge, MA, United States, Jackie Baek, Thodoris Lykouris

We study a model of social learning from reviews where customers are computationally limited and make purchases based on reading only the first few reviews displayed by the platform. Under this bounded rationality, we establish that the review ordering policy can have a significant impact. In particular, the popular Newest First ordering induces a negative review to persist as the most recent review longer than a positive review. This phenomenon, which we term the "Cost of Newest First", can make the long-term revenue unboundedly lower than a counterpart where reviews are exogenously drawn for each customer.

We show that the impact of the "Cost of Newest First" can be mitigated under dynamic pricing, which allows prices to depend on the set of displayed reviews. Under the optimal dynamic pricing policy, the revenue loss is at most a factor of 2. On the way, we identify a structural property for this optimal dynamic pricing: prices should ensure that the probability of a purchase is the same, regardless of the state of reviews. We also extend this to a model where customers put more weight on more recent reviews (and discount older reviews based on their time of posting). We show that Newest First is still not the optimal ordering policy if customers discount slowly.

Lastly, we corroborate our theoretical findings using real-world review datasets. In line with our theoretical results, we find that the average rating of the first page of reviews is statistically significantly smaller than the overall average rating.

3 - Measuring User Strategization on Recommender Systems

Andrew Ilyas, Stanford Statistics, Palo Alto, CA, United States, Sarah Cen, Jennifer Allen, Hannah Li, Aleksander Madry

Modern recommendation algorithms are data-driven: they generate personalized recommendations by observing users' past behaviors. A common assumption in recommendation is that how a user interacts with a piece of content (e.g., whether they choose to "like" it) is a reflection of the content, but not of the algorithm that generated it. Although this assumption is convenient, it fails to capture user strategization: that users may attempt to shape their future recommendations by adapting their behavior to the recommendation algorithm. In this work, we test for user strategization using a lab experiment and survey. We adopt a model of strategization in which strategic users select their engagement behavior based not only on the content, but also on how their behavior affects downstream recommendations.

Using a custom music player, we study how users respond to different information about their recommendation algorithm as well as to different incentives about how their actions affect downstream outcomes. We find strong evidence of strategization across outcome metrics, including participants' dwell time and use of "likes." For example, users who are told their behavior will be aggregated behave very differently from users who are told their behavior will be used to generate recommendations. In our post-experiment survey, nearly half of participants self-report strategizing "in the wild," with some stating that they ignore content they actually like to avoid over-recommendation of that content in the future. Together, our findings suggest that user strategization is common and that platforms cannot ignore the effect of their algorithms on user behavior.

4 - Balancing Producer Fairness and Efficiency via Prior-Weighted Rating System Design

Thomas Ma, MIT Sloan School of Management, Cambridge, MA, United States, Michael Bernstein, Ramesh Johari, Nikhil Garg

Online marketplaces use rating systems to promote the discovery of high-quality products. However, these systems also lead to high variance in producers' economic outcomes: a new producer who sells high-quality items, may unluckily receive a low rating early, severely impacting their future popularity. We investigate the design of rating systems that balance the goals of identifying high-quality products (``efficiency'') and minimizing the variance in outcomes of producers of similar quality (individual ``producer fairness'').

We show that there is a trade-off between these two goals: systems that promote efficiency are necessarily less individually fair to producers. We introduce prior-weighted rating systems as an approach to managing this trade-off. Informally, the system we propose sets a system-wide prior for the quality of an incoming product; subsequently, the system updates that prior to a posterior for product quality based on user-generated ratings over time. We show theoretically that in markets where products accrue reviews at an equal rate, a stronger prior makes the

marketplace discount early ratings data (increasing individual fairness), but slows the platform in learning about true item quality (hampering efficiency). We further analyze this trade-off in a responsive market where customers make decisions based on historical ratings. Through calibrated simulations using 19 different real-world datasets, we show that the choice of prior strength mediates the same trade-off in this setting. We demonstrate that by tuning the prior as a design choice in a prior-weighted rating system, platforms can be intentional about the balance between efficiency and producer fairness.

MC10

Summit - 330

Learning Augmented Mechanism Design

Invited Session

Auctions and Market Design

Chair: Vasilis Gkatzelis, Drexel University, Philadelphia, 19104

Co-Chair: Xizhi Tan

1 - Strategic Facility Location with Predictions

Vasilis Gkatzelis, Drexel University, Philadelphia, PA, United States, Priyank Agrawal, Eric Balkanski, Tingting Ou, Xizhi Tan

We introduce an alternative model for the design and analysis of strategyproof mechanisms that is motivated by the recent surge of work in “learning-augmented algorithms”. Aiming to complement the traditional approach in computer science, which analyzes the performance of algorithms based on worst-case instances, this line of work has focused on the design and analysis of algorithms that are enhanced with machine-learned predictions regarding the optimal solution. The algorithms can use the predictions as a guide to inform their decisions, and the goal is to achieve much stronger performance guarantees when these predictions are accurate (consistency), while also maintaining near-optimal worst-case guarantees, even if these predictions are very inaccurate (robustness).

We initiate the design and analysis of strategyproof mechanisms that are augmented with predictions regarding the private information of the participating agents. To exhibit the important benefits of this approach, we revisit the canonical problem of facility location with strategic agents in the two-dimensional Euclidean space. We study both the egalitarian and utilitarian social cost functions, and we propose new strategyproof mechanisms that leverage predictions to guarantee an optimal trade-off between consistency and robustness guarantees. This provides the designer with a menu of mechanism options to choose from, depending on her confidence regarding the prediction accuracy. Furthermore, we also prove parameterized approximation results as a function of the prediction error, showing that our mechanisms perform well even when the predictions are not fully accurate.

2 - MAC Advice for Facility Location Mechanism Design

Zohar Barak, Tel Aviv University, Tel Aviv, Israel, Anupam Gupta, Inbal Talgam-Cohen

Algorithms with predictions have attracted much attention in the last years. We study the k-facility location mechanism design problem with predictions,
where the n agents are strategic and might misreport their location.
Unlike previous models, where predictions are for the k optimal facility locations, we receive n predictions for the locations of each of the agents.
However, these predictions are only "mostly" and "approximately" correct (or MAC for short) --- i.e., some fraction of the predicted locations are allowed to be arbitrarily incorrect,
and the remainder of the predictions are allowed to be correct up to an epsilon-error.
We make no assumption on the independence of the errors. Can such predictions allow us to beat the current best bounds for strategyproof facility location?

We show that the geometric median of a set of points is naturally robust under corruptions, which leads to an algorithm
for single-facility location with MAC predictions. We extend the robustness result to a "balanced" variant of the k facilities case.
Without balancedness, we show that robustness completely breaks down, even for k=2 on a line.
For this "unbalanced" setting, we devise a truthful random mechanism that outperforms the best known result of Lu et al.~[2010], which does not use predictions.
En route, we introduce the problem of "second" facility location (when the first facility's location is already fixed).
Our robustness findings of the 1-median, and more generally k-medians, may be of independent interest as quantitative versions of classic breakdown-point results in robust statistics and robust clustering.

3 - Learning-Augmented Metric Distortion via (p,q)-Veto Core

Xizhi Tan, Drexel University, Philadelphia, PA, United States, Ben Berger, Michal Feldman, Vasilis Gkatzelis

In the metric distortion problem, there is a set of candidates and a set of voters, all residing in the same metric space. The objective is to choose a candidate with the minimum social cost, defined as the total distance of the chosen candidate from all voters. The challenge is that the algorithm receives only ordinal input from each voter, in the form of a ranked list of candidates in non-decreasing order of their distances from the voter, whereas the objective function is cardinal. The distortion of an algorithm is its worst-case approximation factor with respect to the optimal social cost. A series of papers culminated in a 3-distortion algorithm, which is tight with respect to all deterministic algorithms. Aiming to overcome the limitations of worst-case analysis, we revisit the metric distortion problem through the learning-augmented framework, where the algorithm is provided with some (machine-learned) prediction regarding the optimal candidate. The quality of this prediction is unknown, and the goal is to evaluate the performance of the algorithm under a perfectly accurate prediction (known as consistency), while simultaneously providing worst-case guarantees even for arbitrarily inaccurate predictions (known as robustness). This talk we will fully characterize the robustness-consistency Pareto frontier for the metric distortion problem. We will also discuss some new tight bounds for metric distortion without predictions that helped in proving the tight guarantees, which is of individual interest.

4 - Strategyproof Scheduling with Predictions

Eric Balkanski, Columbia University, New York, NY, United States, Vasilis Gkatzelis, Xizhi Tan

In their seminal paper that initiated the field of algorithmic mechanism design, Nisan and Ronen studied the problem of designing strategyproof mechanisms for scheduling jobs on unrelated machines aiming to minimize the makespan. They provided a strategyproof mechanism that achieves an n-approximation and made the bold conjecture that this is the best approximation achievable by any deterministic strategyproof scheduling mechanism. After more than two decades and several efforts, Christodoulou et al. have recently validated the conjecture. This strong negative result, however, heavily depends on the fact that the performance of these mechanisms is evaluated using worst-case analysis. To overcome such overly pessimistic, and often uninformative, worst-case bounds, a surge of recent work has focused on the ``learning-augmented framework", whose goal is to leverage machine-learned predictions to obtain improved approximations when these predictions are accurate (consistency), while also achieving near-optimal worst-case approximations even when the predictions are arbitrarily wrong (robustness).

In this work, we study the classic strategic scheduling problem of Nisan and Ronen using the learning-augmented framework and give a deterministic polynomial-time strategyproof mechanism that is 6-consistent and 2n-robust. We thus achieve the ``best of both worlds": an O(1) consistency and an O(n) robustness that asymptotically matches the best approximation possible. We then extend this result to provide more general worst-case approximation guarantees as a function of the prediction error. Finally, we complement our positive results by showing that any 1-consistent deterministic strategyproof mechanism has unbounded robustness.

MC11

Summit - 331

AI, Collaboration, and IT Procurement Innovations

Contributed Session

Chair: Ecenur Oğuz

1 - Collaborative Efforts with Users: A Novel Cloud Service Business Model

Aijun Lu, College of Management and Economics, Tianjin University, Tianjin, China, People's Republic of, Fuzan Chen, Harris Wu, Minqiang Li, Haiyang Feng

It was once believed that Cloud Service Providers (CSPs) primarily held responsibility for service availability, yet user efforts are equally essential. Some leading CSPs like Amazon have shifted towards a collaborative approach by implementing the Customer Enablement (CE) system. It comprises resources, tools, and expertise programs to enhance users' cloud capabilities, which suggests a transformative business model promoting joint efforts in the cloud industry.

Inspired by such observations, this paper explores several questions: (1) How should a CSP optimally allocate investments between availability and the CE system? (2) What are users' optimal availability efforts and CE-based efforts? (3) How does the unit downtime loss, affect the efforts interactions between the two parties? (4) Are the service quality improvements effective in increasing the CSP's profitability? Employing game-theoretic models, we derive the CSP's optimal pricing, availability efforts, and CE function level, as well as users' optimal efforts. We further analyze the equilibrium results to answer the above questions and investigate some management insights.

This study reveals several interesting findings. First, with increasing unit downtime loss, the CSP prefers enhancing the CE system over increasing its availability efforts, which leads to an effective improvement in service quality due to strengthened CSP-user collaboration. Second, the CSP's profits first decrease and then increase as downtime loss grows, illustrating that higher quality from intensified efforts does not always translate into higher returns. This paper has strong management implications for the cloud industry, and the findings provide valuable insights for the CSPs.

2 - Enhancing Success in IT Procurement within State Agencies

Mackey Cardinell, Western Washington University, Bellingham, WA, United States, Stella Hua

In this study, we investigate how state agencies' commitment to Lean practices and IT vendors' agile development approaches enhance the success of IT procurement. We examine IT projects within the California Department of Motor Vehicles (DMV) and other state agencies. We look into the impact of agile development practices such as minimum viable products (MVPs) and proof of concepts (POCs) on vendor selection, evaluation, and progress monitoring. Our findings contribute to the procurement literature and provide insights for state agencies pursuing IT modernization.

3 - Appointment Management from Theory to Application: 'Qararment' An Innovative Decision Support System

Majed Hadid, Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar, Adel Elomri

Background: Appointment management in the service industry confronts challenges due to subprocesses' stochastic nature and interdependence. The gap between theoretical models and practical application is wide, particularly for end-users without operations research background. This situation is compounded by the prevalence of commercial software that prioritizes digitization over optimization, leaving a void for optimization-based solutions.

Aim: Our research develops Qararment, a machine learning and simulation-based optimization software for appointment planning and scheduling, adaptable to various process configurations and management policies, addressing multiple objectives.

Methodology: We developed a self-configuring, data-driven simulation model of the service process, synergistically combined with genetic algorithms to constitute the core engine of the software. This combination allows for a dynamic and flexible adaptation to different service settings, making the software highly versatile.

Results: Implemented in a major chemotherapy center, Qararment showcased significant adaptability, improving patient waiting times, staff overtime, and computational efficiency by 37%, 17%, and 48%, respectively. The user-friendly interface facilitated model configuration and optimal scheduling, receiving positive feedback for ease of use and effectiveness.

Conclusions: This research highlights the challenges and opportunities of implementing operations research and optimization-oriented solutions in real-world environments, considering the perspectives of end-users. It opens new avenues for exploring the practical application of service operations management theories, particularly in environments that have not been fully investigated, emphasizing the need for solutions that are both sophisticated in capability and simple in user experience.

4 - Point Cloud Registration With Quantile Assignment

Ecenur Oğuz, Northwestern University, Evanston, IL, United States, Yalim Dogan, Ugur Gudukbay, Oya Karasan, Mustafa Pinar

Point cloud registration is a fundamental problem in computer vision. The problem encompasses critical tasks such as feature estimation, correspondence matching, and transformation estimation. The point cloud registration problem can be cast as a Quantile Assignment problem. We refined the Quantile Assignment algorithm by integrating prevalent feature descriptors and transformation estimation methods to enhance the correspondence between the source and target point clouds. We evaluated the performances of these descriptors and methods with our approach through controlled experiments on a dataset we constructed using well-known 3D models. This systematic investigation led us to identify the most suitable methods for complementing our approach. Subsequently, we devised a new end-to-end, coarse-to-fine pairwise point cloud registration framework. Finally, we tested our framework on indoor and outdoor benchmark datasets and compared our results with state-of-the-art point cloud registration methods.

MC12

Summit - 332

Optimizing Networked Environments and Demand Matching

Flash Session

1 - Enhancing Communication Network Traffic Engineering through Machine Learning

Sharon Arroyo, Boeing Company, Seattle, WA, United States, Yingxi li, Raju Mattikalli, Ali Teshnizi, Madeleine Udell, Mikel Woo

Tactical communication networks currently struggle to maintain throughput and service levels to changing network conditions and traffic. Software defined networking (SDN) offers a solution by allowing for greater control and optimization of network capacity, resulting in improved performance in terms of throughput, delay and loss. One key aspect of SDN is traffic engineering (TE).

TE in tactical SDN involves computing routes for point-to-multi-point service requests within a known communications network. This network is defined by point-to-point links, each with its own bandwidth and delay characteristics. Service requests are characterized by priority, bandwidth consumption, and maximum allowed delay.

Traditional mathematical programming methods, like mixed integer optimization solvers, have limitations in solving real-world traffic engineering (TE) problems. These methods face difficulties in handling the complex requirements and scalability demands of TE, particularly under real-time constraints. To address these challenges, we have turned to reinforcement learning and supervised learning algorithms. Our approach has yielded high-quality solutions significantly faster than traditional TE methods. In our presentation, we will showcase the results and provide valuable insights into effectively tackling hard constraints in combinatorial optimization problems.

Our goal is to bridge the gap between traditional mathematical programming and machine learning approaches. By harnessing the strengths of both fields, we can effectively tackle the intricacies of TE problems and meet the real-time demands of communication networks

2 - A Resource Allocation Game Between Attackers and Defenders in a Multi-Entrance, Multi-Layered, and Multi-Target System

Ting Zan, University at Buffalo, Buffalo, NY, United States, Jun Zhuang

Many security systems, including airports, schools, and public transit, feature multiple entrances and interconnected layers to ensure efficacy and safety. The optimal allocation of defensive resources across these layers to minimize overall system risk, especially against a strategic attacker, has been underexplored in the literature. This research addresses this gap by examining a sequential game-theoretical resource allocation model and demonstrating its results through a case study.

3 - Junction Hub Location Problem

CHANCHAL KUMAR SALODE, Indian Institute of Technology Delhi, New Delhi, India, Prasanna Ramamoorthy

In this study, we introduce the junction hub location problem, focusing on locating a set of interconnected hub lines, where hubs are connected by means of a line, while assigning non-hub nodes to hubs in a single allocation scheme. These hub lines intersect at a common hub referred to as junctions. This problem finds applications in various domains such as public transportation, telecommunications, and rapid transit systems, where the cost of designing networks is high compared to routing costs, making full interconnection of hubs impractical. The objective is to minimize transportation costs between all pairs of origin and destination nodes. To address this problem, we propose a Mixed Integer Linear Programming (MILP) model. Additionally, to solve this problem on large scale, we propose Benders branch and cut algorithm.

Computational results obtained on two sets of benchmark instances confirm the efficiency of the proposed algorithm, which is considerably faster and able to solve larger instances than a general purpose solver.

4 - Using combinatorial auctions to allocate parcel logistic services in Hyperconnected City Logistics

Simon Kwon, Georgia Institute of Technology, Atlanta, GA, United States, Walid Klibi, Mathieu Dahan, Benoit Montreuil

Aligned with Physical Internet-enabled hyperconnected city logistics initiatives, this study explores a setting in which an urban logistics orchestrator aims to design a multi-party service network. In this network, logistic activities, such as hub processing and transportation, are assigned to specific service providers through term contracts via a single-round combinatorial auction. The orchestrator's role is to cater to citywide origin-destination demand flows, ensuring delivery time guarantees and service reliability while maximizing profit. The path of these flows is segmented into logistic activities, including transport and hub processing. Each logistic activity is associated with a reliable service level agreement (e.g., 30 minutes for transport in Local cell 1 at 99.9%) to fulfill origin-destination time guarantees, and service providers participating in the auction (bidders) must adhere to their claimed service level agreements.

Due to the decision lag caused by the auction mechanism's tactical nature, bidders may not uphold their claimed service level agreements after auction, leading to uncertainty in origin-destination services and potential profit erosion. Additionally, considering the historical service performance of bidders, different risk levels for the orchestrator result in varying reliability in bidders. Therefore, the orchestrator must incorporate these uncertainties when determining winners to achieve desired reliability in origin-destination services.

We present a chance-constrained stochastic mixed-integer programming optimization model for this problem, utilizing chance constraints to ensure origin-destination deliveries at desired reliability levels while maximizing expected profit. Our numerical results demonstrate the impact of the chance-constrained programming approach on delivery guarantees, contrasting with the traditional deterministic approach for origin-destination flows.

5 - Traffic control as a process: A novel outlook on task execution and operational performance

Léon Sobrie, Virginia Tech, Falls Church, VA, United States, Jasmijn De Clercq, Bart Roets, Konstantinos Triantis

This paper proposes a novel outlook toward traffic control by centralizing the operator and the tasks they execute. The operator's actions are modeled as a process, consistent with a functional representation of a system, in which we unravel sequences of different task types. The analysis delves into multitasking and the time lag between events as two explanatory characteristics for the operational performance of traffic control. The method is operationalized and validated at Infrabel, Belgium's national railway infrastructure company. The setup is generalized for other traffic control settings and more general settings where task execution can be modeled as a process. Also, linkages between traffic control as a process and different modeling paradigms (such as machine learning, DEA and systems thinking) are discussed.

6 - Braess Paradoxes in Coupled Power and Transportation Systems

Minghao Mou, Purdue University, West Lafayette, IN, United States, Junjie Qin

Transportation electrification introduces strong coupling between the power and transportation systems. In this paper, we generalize the classical notion of Braess Paradox to coupled power and transportation systems, and examine how the cross-system coupling induces new types of Braess Paradoxes. To this end, we model the power and transportation networks as graphs, coupled with charging points connecting to nodes in both graphs. The power system operation is characterized by the economic dispatch optimization, while the transportation system user equilibrium models travelers' route and charging choices. By analyzing simple coupled systems, we demonstrate that capacity expansion in either transportation or power system can deteriorate the performance of both systems, and uncover the fundamental causes for such new Braess Paradoxes, leading to managerial insights for infrastructure planners. For general networks, we develop efficient algorithms for computing the equilibrium of and identifying Braess Paradoxes for the coupled system.

7 - Smarter Traffic Management: Leveraging Deep Learning for Enhanced Flow Predictions

Zhe Fu, UC Berkeley, Berkeley, CA, United States

Effective traffic management is essential for fostering smarter, sustainable cities. In this study, we revolutionize traffic flow predictions by applying deep learning to the Lighthill-Whitham-Richards (LWR) traffic model, a foundational macroscopic framework governed by hyperbolic partial differential equations (PDEs). Traditional numerical methods for solving these PDEs lack the flexibility and computational efficiency required for real-time urban traffic management, often resulting in suboptimal decision-making.

Our innovative approach employs deep learning neural networks as advanced numerical approximators, integrating traditional numerical principles with the governing conservation laws of traffic flow. This methodology significantly enhances predictive accuracy and computational efficiency in traffic simulations, leading to more informed and agile traffic management decisions. The benefits of our approach include:
1) Enhanced prediction accuracy for traffic flow, leading to more reliable estimates of travel times and expected time of arrival (ETA), crucial for route planning and congestion management.
2) Reduced computational complexity, facilitating quicker adjustments in traffic control systems to accommodate real-time conditions.
3) Improved adaptability across diverse traffic models and flux functions, increasing the resilience of traffic systems against varying urban dynamics.

By advancing deep learning applications in traffic flow modeling, our research supports smarter, data-driven decisions for better world outcomes, optimizing urban mobility and reducing environmental impacts through enhanced traffic management strategies.

MC13

Summit - 333

Advanced Analytics in Health and Well-being

Contributed Session

Chair: Mohammad Rahimi, Virginia Tech, Blacksburg, VA, United States

1 - Enhancing Depression Detection in Social Media Texts Using Generative Adversarial Networks

Kaung-Yu Chao, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Puli, Taiwan, Jing-Rung Yu

In contemporary society, the prevalence of depression is rising, leading to many negative outcomes. Research indicates that mental health status can be inferred from users' texts, images, and interaction patterns on social media. Early identification of mental health issues through social media text analysis is crucial for timely intervention and treatment. However, obtaining large labeled depression datasets is challenging due to the intensive labor required. Therefore, we use datasets from Kaggle and employ Generative Adversarial Networks (GANs) to augment the data for our training and testing datasets. To enhance the robustness of the BERT model in detecting depression in texts, we utilize various loss functions, such as cross-entropy. We evaluate whether using GANs or adjusting the loss function helps maintain or improve the model's accuracy and robustness. This approach aims to strengthen sentiment analysis models based on the BERT model by using GANs and modifying the loss function. We compare the performance of various BERT models with and without using GANs and adjusting the loss function.

2 - Enhancing Sleep Apnea Diagnosis with Machine Learning: A Study on Wearable Biometrics and Smartphone Application Data

HyungJu Kim, Korea University, Seoul, Korea, Republic of, Taesu Cheong, Chul-Hyun Cho

Sleep apnea is a condition in which breathing stops temporarily during sleep. It causes irregular breathing patterns and reduced oxygen levels, which can lead to poor quality of rest and can cause fatigue and difficulty concentrating during the day. There are many causes of sleep apnea, including obesity, age, smoking, and alcohol. Sleep apnea can be diagnosed through polysomnography to determine the presence and severity of apnea. Treatment can also be provided through questionnaires and patient information. However, in mild cases, it is difficult for patients to recognize the disease, and questionnaires can be difficult to diagnose and treat due to the subjective nature of the patient. There is a need for an objective and quantitative method to complement this. We aimed to classify sleep apnea severity with a machine learning model using biometric data from wearable devices and self-reported data through a smartphone application. Due to the small number of patients, we applied oversampling technique, and the classification model used tree-based models such as ngboost and XGB, and performed parameter tuning. The results showed significant classification performance (AUC 0.70 ~ 0.80). We believe that this methodology can complement the existing diagnostic process.

3 - A New Genomics-based Psychographic Segmentation Strategy

Camelia Taheri Protzel, University of North Carolina - Charlotte, Charlotte, NC, United States

Advances in behavioral genetics provide a paradigm shift in the development of a genomics-based psychographic segmentation strategy. Genetics can explain most of the systematic variation between individuals, the continuity of behavioral and personality traits, as well as 50% of the variance in human traits. Leveraging that all human behaviors and psychological traits are influenced in some way by the individual's genetic constitution, a theoretical framework is presented for the definition of a new genomics-based psychographic segmentation. Genomics-based psychographic segmentation takes into account the genetic predisposition measured by polygenic scores (PGS), a combination of thousands of single-nucleotide polymorphisms (SNPs) associated with a specific trait that can estimate heritability and predict phenotypes. The study empirically shows the applicability of this new genomics-based psychographic segmentation through K-mean clustering analyses of the alcohol consumption market using 7 different PGS related to personality and cognitive traits. The study increases the predictive power of consumer behavior and marketing segmentation leveraging molecular genetics and 150 years of behavioral genetics replicable findings. It presents for the first-time fundamental principles from behavioral genetics to lay the ground for genomics marketing and the transformation of segmentation strategies. It proposes the segmentation of markets through the genetic propensity of consumers. It not only highlights embryonic research in genomics marketing but also shows the practical application of genomics segmentation through the usage of molecular genetics to create clusters and understand consumption patterns of each subset.

4 - Optimizing Postpartum Depression Screening Systems in the Neonatal Intensive Care Unit and Pediatric Emergency Department: A Simulation-based Analysis of the Workflow

Fatima Sadjadpour, Virginia Tech, Falls Church, VA, United States, Niyousha Hosseinichimeh, Lenore Jarvis, Sofia Perazzo, Lamia Soghier

Perinatal mood and anxiety disorders (PMAD) are a large group of disorders that encompass postpartum depression, anxiety, acute stress disorder, perinatal psychosis, obsessive-compulsive disorder, and post-traumatic stress disorder. Postpartum depression, occurring within the first year after childbirth, affects 13-15% of mothers globally, posing a significant obstetric complication and a prominent factor in maternal suicide. The impact of postpartum depression extends beyond mothers to affect all caregivers, and the long-term consequences result in higher rates of anxiety, depression, and behavioral issues among infants. Particularly vulnerable are caregivers of infants in neonatal intensive care units (NICU), where postpartum depression prevalence is markedly higher around 39-45%, leading to recurrent visits to the pediatric emergency department (PED). Despite its pervasive impact, clinical recognition, and treatment of postpartum depression in these settings remain challenging. At Children's National Hospital in 2021, internal benchmarking data shows that only 64% (n=577) of eligible NICU parents and 34% (n=1270) of eligible PED parents are screened even with an established PMAD team of social workers and screeners. To address these complexities, a system dynamics group model building (GMB) approach is employed to analyze the current screening system and develop a simulation model aimed at enhancing screening and referral processes in NICU and PED settings. Findings from GMB sessions included interventions, such as 24/7 access and staffing, standardizing screening as part of routine care, and improving public relations. By integrating these solutions into the system dynamics model, we aim to test their potential impact within the broader system.

5 - Which is more beneficial, superficial or deep integration? Exploring the impact of work-family integration on employees' creative process engagement

Wei Liu, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Ning Wang, Mi Zhou, Meng Qin

With the dual influence of digital human resources and the post-epidemic period, remote work has emerged as a prevailing trend. It is inevitable that work-family integration will occur. Existing literature suggests that work-family integration can take multiple forms. Hence, we classified work-family integration as superficial integration and deep integration according to content dimensions. Recognizing the importance of not only the outcomes but also the process of innovation, we proposed a model to examine how different forms of work-family integration affect creative process engagement. We hypothesized that work-related rumination (affective rumination and problem-solving pondering) would mediate these relationships, with the partner's communal motivation acting as a moderating factor. In study 1, using a within-person daily diary study lasting ten days (data was collected from 21 dual-earner couples), we validated our proposed model. In study 2, we also used a between-person design from 228 employees to determine whether our model held. As predicted, superficial or deep integration affects creative process engagement positively. What's more, problem-solving pondering mediates deep integration to creative process engagement, while affective rumination does not mediate superficial integration to creative process engagement. Additionally, partner's communal motivation would buffer the indirect effect of deep integration on creative process engagement through problem-solving pondering, while partner's communal motivation would not buffer the indirect effect of superficial integration on creative process engagement through affective rumination. Overall, this study advances separating work-family integration into superficial and deep types, demonstrating that creative process engagement follows a dual path through these two types.

6 - What Do You Mean You&Rsquo;Re Sorry?! You&Rsquo;Re Just a Machine! Empathy in Conversational AI: the Case of Large Language Models in Mental Health

Mohammad Rahimi, Virginia Tech, Blacksburg, VA, United States, Idris Adjerid, Paul Lowry, Warren Bickel

The mental health crisis has been rapidly growing over the last few decades and further exacerbated by the COVID-19 pandemic, turning it into a deadly global challenge that leads to the loss of twelve billion working days each year, accounts for more than 14% of all deaths, and will cost the world's economy $16 trillion by 2030. Half of the world's population will experience a mental health disorder during their lifetime. A major problem is the lack of qualified personnel to address this crisis. We explore whether AI agents, with their latest advancements, can be a viable solution. Emotional support is a key factor in this domain as it encourages clients' information self-disclosure, which is usually the first step when seeking help in this context. Recent studies indicate that modern AI agents can emulate human empathy with high precision. Such advancements hold promise, especially in fields like mental health where empathy and comprehension are paramount. Past studies have either focused solely on emulating human-like empathy or treated empathy as a simple and aggregated concept. Our study is the first to address these two important gaps in the literature, offering a granular examination of AI expressing components of human empathy (e.g., cognitive, and affective) and evaluating the efficacy of complimenting human-like empathy with unique AI data capabilities–a new phenomenon we coin as *collective artificial empathy*. We will conduct a 3x2 randomized between-subjects experiment with three levels of empathy (low, medium, high) and two different types of empathy (human-like and machine-like empathy).

MC14

Summit - 334

Machine Learning in Finance I

Invited Session

Finance

Chair: Markus Pelger, Stanford University, Stanford, CA, United States

1 - Understanding Economic Decision Making with Machine Learning

Lea Stern, University of Washington, Seattle, WA, United States

upcoming

2 - A Study Of The Reliability Of Crypto Data Provision

Gustavo Schwenkler, Santa Clara University, Santa Clara, CA, United States, Aakash Shah, Darren Yang

We analyze the supply of cryptocurrency data from leading providers and uncover significant quality issues. We document the impact of these issues on the empirical analysis of cryptocurrency markets, emphasizing the risks of relying on single sources. We propose a methodology to aggregate data across providers that ensures consistency and demonstrate its benefits in an asset pricing study. We recommend the adoption of a unified identification system for cryptocurrencies, akin to traditional financial markets, to enhance data reliability. We also flag a need for potential regulatory oversight to ensure consumer protection in the market for crypto data.

3 - Long-History PCA in a Dynamic Factor Model with Weaker Loadings

Baeho Kim, Korea University Business School, Seoul, Korea, Republic of, Robert Anderson, Dean Ryu

Estimated covariance and precision matrices of asset returns significantly influence the set of portfolios compliant with risk budgets and their potential losses. Statistical risk modeling approaches often assume temporal stability for consistency with a static factor structure, typically estimated within $T_M \sim 250$ days of data history, resulting in finite-sample estimation error when the dimension of the population exceeds the number of observations. Our study introduces Long-History PCA (LH-PCA), utilizing extended data histories (e.g., $T_L \sim 1500$ trading days) to forecast the daily risk profile based on dynamic factor structures with heterogeneous factor strengths. The use of a longer data history mitigates excess dispersion bias in the estimated factor loadings, particularly in the presence of weaker factors. Augmented by Responsive

Covariance Adjustment (RCA) employing a short half-life (T_S) of 40 days, our approach substantially mitigates second-order risk biases compared to traditional approaches using medium horizons (T_M) of 250 days, in both simulations and empirical studies.

4 - Do AI-Algorithmic Traders Lead to Market Instability? A Multi-Agent Reinforcement Learning Approach

Yang Fan, Stanford University, Stanford, CA, United States, Markus Pelger, Xintong Yu

We show that AI-powered algorithmic trading can lead to inefficient and instable market. We investigate the impact of algorithmic traders on market equilibrium using a multi-agent deep reinforcement learning framework. We develop a conceptual simulation framework to study the market equilibrium among informed speculators with asymmetric information under different market conditions. Our findings reveal that more informed algorithmic traders engage in manipulative behavior, leading to local market price bubbles and deviations from the fundamental price. The informed AI-traders exploit and actively distort the learning process of less informed AI-traders. These results are robust across different number of agents and buying power scenarios. Unlike existing studies on algorithmic collusion, our research highlights the novel finding of the emergence of local price bubbles. Given the rapid growth in AI-powered algorithmic trading, our results identify a new form of systematic risk.

MC15

Summit - 335

Sequential Learning Problems in Operations Management

Invited Session

Revenue Management and Pricing

Chair: Dongwook Shin, HKUST Business School, Clear Water Bay, Hong Kong

Co-Chair: Dohyun Ahn, CUHK, Shatin, Hong Kong

1 - On the Optimality of Bisection Search in Noisy Environments

Sanghwa Kim, KAIST, Daejeon, Korea, Republic of, Dohyun Ahn, Seungki Min

In adaptive testing, the tester sequentially determines the difficulty of the next query in order to identify the ability level of the testee as efficiently as possible. We study a stochastic version of the bisection search algorithm, a simple heuristic that selects the next difficulty level at which a positive outcome and a negative outcome become equally likely with respect to the currently estimated parameter. Adopting bandit and experimental design approaches, we analyze the optimality of this algorithm under three criteria: (i) sample complexity in the fixed confidence setting, (ii) error rate in the fixed budget setting, and (iii) Fisher information. For each criteria, we provide a theoretical limit that no algorithm can achieve, and provide sufficient conditions under which this algorithm achieves the limit asymptotically (up to some constant factor).

2 - Mixed-Effects Contextual Bandits

Kyungbok Lee, Seoul National University, Seoul, Korea, Republic of, Myunghee Cho Paik, Min-hwan Oh, Gi-Soo Kim

We study a novel variant of a contextual bandit problem with multi-dimensional reward feedback formulated as a mixed-effects model, where the correlations between multiple feedback are induced by sharing stochastic coefficients called random effects. We propose a novel algorithm, Mixed-Effects Contextual UCB (ME-CUCB), achieving tildeO(d sqrt(mT)) regret bound after T rounds where d is the dimension of contexts and m is the dimension of outcomes, with either known or unknown covariance structure. This is a tighter regret bound than that of the naive canonical linear bandit algorithm ignoring the correlations among rewards. We prove a lower bound of Omega(d sqrt(mT)) matching the upper bound up to logarithmic factors. To our knowledge, this is the first work providing a regret analysis for mixed-effects models and algorithms involving weighted least-squares estimators. Our theoretical analysis faces a significant technical challenge in that the error terms do not constitute martingales since the weights depend on the rewards. We overcome this challenge by using covering numbers, of theoretical interest in its own right. We provide numerical experiments demonstrating the advantage of our proposed algorithm, supporting the theoretical claims.

3 - Simple Data-Driven Solutions and Uncertainty Quantification for Multistage Stochastic Optimization

Yunhao Yan, Columbia University, New York, NY, United States, Henry Lam

Multistage stochastic optimization problems appear commonly in various disciplines in operations management and system control. While computational methods have been studied, they are mostly restricted to two-stage problems due to the exponentially increasing complexity arising from multistage scenarios. We argue that for some interesting applications, these problems exhibit stagewise independence and, with the use of suitable classes of policies, they are amenable to simple data-driven solutions and uncertainty quantification methodologies based on natural multisample generalizations of the well-known sample average approximation (SAA) and single replication procedure. We justify the statistical consistency and coverage guarantees of our approaches, and demonstrate how they reduce variances compared to the naive use of SAA. Our results require developing several functional limit theorems of the so-called multisample U-processes that closely connect to multistage stochastic optimization, which allow us to work with closed-loop control policies beyond Euclidean parametrization classes.

4 - Collaborative Intelligence in Sequential Experiments: A Human-in-the-Loop Framework for Drug Discovery

Jinghai He, University of California, Berkeley, Berkeley, CA, United States, Cheng Hua, Yingfei Wang, Zeyu Zheng

Drug discovery is a complex process that involves sequentially screening and examining a vast array of molecules to identify those with the target properties. This process, also referred to as sequential experimentation, faces challenges due to the vast search space, the rarity of target molecules, and constraints imposed by limited data and experimental budgets. To address these challenges, we introduce a human-in-the-loop framework for sequential experiments in drug discovery. This collaborative approach combines human expert knowledge with deep learning algorithms, enhancing the discovery of target molecules within a specified experimental budget. The proposed algorithm processes experimental data to recommend both promising molecules and those that could improve its performance to human experts. Human experts retain the final decision-making authority based on these recommendations and their domain expertise, including the ability to override algorithmic recommendations. We applied our method to drug discovery tasks using real-world data and found that it consistently outperforms all baseline methods, including those which rely solely on human or algorithmic input. This demonstrates the complementarity between human experts and the algorithm. Our results provide key insights into the levels of humans' domain knowledge, the importance of meta-knowledge, and effective work delegation strategies. Our findings suggest that such a framework can significantly accelerate the development of new vaccines and drugs by leveraging the best of both human and artificial intelligence.

MC16

Summit - 336

Frontiers of Sequential Optimization: Learning, Algorithmic Advice, and Information Design

Invited Session

Revenue Management and Pricing

Chair: Anand Kalvit, Stanford University, Stanford, CA, United States

Co-Chair: Yonathan Gur, Stanford University, Netflix, Stanford, CA, United States

1 - Just a Few Bits: Improving Human Decision-Making with Minimal Algorithmic Intervention

Stefanos Poulidis, INSEAD, Fontainebleau, France, Haosen Ge, Hamsa Bastani, Osbert Bastani

Interpretable algorithmic advice has been shown to enhance human decision-making and performance. However, current forms of algorithmic advice predominantly provide action signals, which prescribe specific courses of action. These action signals can be challenging to estimate, brittle, and susceptible to unobserved confounders. In this work, we introduce attention signals as an alternative. Attention signals are simple binary indicators that do not recommend specific actions. We develop a novel algorithm to determine the optimal timing for offering action and attention algorithmic signals within the context of chess, which can be readily extended to general Markov Decision Processes (MDPs). Then, we conduct a comprehensive behavioral experiment including 36 titled chess masters (e.g., Grandmasters), to study the effects of attention signals on human decision-making and performance. Our findings indicate that attention signals lead to a more efficient distribution of human search effort, more profitable actions, and substantial performance gains. We synthesize our findings to uncover the key mechanism driving the performance gains induced by attention signals, providing insights for the effective deployment of algorithmic advice.

2 - Incentive-Aware Synthetic Control: Accurate Counterfactual Estimation via Incentivized Exploration

Keegan Harris, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, United States, Daniel Ngo, Anish Agarwal, Vasilis Syrgkanis, Steven Wu

We consider the setting of synthetic control methods (SCMs), a canonical approach used to estimate the treatment effect on the treated in a panel data setting. We shed light on a frequently overlooked but ubiquitous assumption made in SCMs of "overlap": a treated unit can be written as some combination---typically, convex or linear combination---of the units that remain under control. We show that if units select their own interventions, and there is sufficiently large heterogeneity between units that prefer different interventions, overlap will not hold. We address this issue by proposing a framework which incentivizes units with different preferences to take interventions they would not normally consider. Specifically, leveraging tools from information design and online learning, we propose a SCM that incentivizes exploration in panel data settings by providing incentive-compatible intervention recommendations to units. We establish this estimator obtains valid counterfactual estimates without the need for an a priori overlap assumption. We extend our results to the setting of synthetic interventions, where the goal is to produce counterfactual outcomes under all interventions, not just control. Finally, we provide two hypothesis tests for determining whether unit overlap holds for a given panel dataset.

3 - Incentivized Exploration via Filtered Posterior Sampling

Yonatan Gur, Stanford University, Netflix, Stanford, CA, United States, Anand Kalvit, Aleksandrs Slivkins

We study the issue of incentivized exploration (IE) in social learning problems where the principal (recommendation algorithm) can leverage information asymmetry to incentivize sequentially-arriving agents to take exploratory actions. This work expands upon extant literature on IE in bandits, and generalizes theory along several practically-relevant dimensions.

4 - Secondary Migration-Aware Placement For Refugees

Salomon Wollenstein-Betech, Stanford University, Stanford, CA, United States, Kostas Bimpikis, Yonathan Gur, Jens Hainmueller

About 17% of refugee migrants worldwide do not stay in their initial destination for more than a year and they seek for a new location -- known as *secondary migration (or onward movements)*-- to achieve self-sufficiency. Current matching methods to allocate refugees to destinations only consider the first destination and do not incorporate explicitly the long-term effect of future migration in their resettlement decisions which restricts the overall success of the match. In this work, we incorporate the notion of secondary migration by formulating and solving an augmented matching problem that takes into account the *long-term* probability of employment and destination capacities. We leverage historical US resettlement data between 2004-2017 to understand the secondary migration patterns and we develop a simulated environment to compare the long-term benefits of using a secondary-migration aware policy against a myopic one. Our results show that our long-term resettlement decisions could reduce in half the number of refugees that experience secondary migration while keeping overall employability at current levels. Moreover, we show that a modest 10% reduction in employability could reduce migration by 40%.

MC17

Summit - 337

Sustainable and Responsible Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Xiao Lei, University of Hong Kong, Hong Kong, N/A, Hong Kong

1 - Dynamic Pricing with Fairness Constraints

Sentao Miao, University of Colorado Boulder, Boulder, CO, United States, Maxime Cohen, Yining Wang

Following the increasing popularity of personalized pricing, there is a growing concern from customers and policy makers regarding fairness considerations. This paper studies the problem of dynamic pricing with unknown demand under two types of fairness constraints: price fairness and demand fairness. For price fairness, the retailer is required to (i) set similar prices for different customer groups (called group fairness) and (ii) ensure that the prices over time for each customer group is relatively stable (called time fairness). We propose an algorithm based on an infrequently-changed upper-confidence-bound (UCB) method, which is proved to yield a near-optimal regret performance. We then leverage this method to address the extension of non-stationary demand, which is particularly relevant for time fairness to prevent price gouging practices. For demand fairness, the retailer is required to satisfy that the resulting demand from different customer groups is relatively similar (e.g., the retailer offers a lower price to students to increase their demand to a similar level as non-students). In this case, we design an algorithm adapted from a primal-dual learning framework and prove that our algorithm also achieves a near-optimal regret performance.

2 - Temporal Fairness in Learning and Earning: Price Protection Guarantee and Phase Transitions

Ruihao Zhu, Cornell University, Ithaca, NY, United States, Qing Feng, Stefanus Jasin

Motivated by the prevalence of "price protection guarantee", which helps to promote temporal fairness in dynamic pricing, we study the impact of such policy on the design of online learning algorithm for data-driven dynamic pricing with initially unknown customer demand. Under the price protection guarantee, a customer who purchased a product in the past can receive a refund from the seller during the so-called price protection period (typically defined as a certain time window after the purchase date) in case the seller decides to lower the price. We consider a setting where a firm sells a product over a horizon of T time steps. For this setting, we characterize how the value of M, the length of price protection period, can affect the optimal regret of the learning process. We show that the optimal regret is $\tilde{\Theta}(\sqrt{T}+\min\{M,\,T^{2/3}\})$. Our results reveal the surprising phase transitions of the optimal regret with respect to M. Specifically, when M is not too large, the optimal regret has no major difference when compared to that of the classic setting with no price protection guarantee. We also show that there exists an upper limit on how much the optimal regret can deteriorate when M grows large.

3 - Heuristic but Effective Pricing with Data Expiry

Jialin Li, Rotman School of Management, Toronto, ON, Canada

Small business owners who cannot afford a data scientist team may resort to data analytics platforms. In compliance with emerging privacy regulations, platform service providers remove data from the database after retaining it for a specified period of time. In this dynamic process, existing demand learning or pricing methods can easily fail in this situation because the platform user can never enjoy a large sample size. In this work, we demonstrate a simple estimator that leads to effective learning and pricing for a firm user who has been setting heuristic prices in history.

4 - Pricing and Control for Digital Addictive Services

Jiacheng Chang, University of Hong Kong, Hong Kong, Hong Kong, Xiao Lei, Feng Tian

Digital addictive services, such as video games and video sharing platforms, have become prevalent in daily life, raising concerns about digital addiction. Unlike traditional addictive goods, these services allow for personalized dynamic pricing and manipulation of experience randomness. In this paper, we investigate optimal pricing strategies for digital service providers and effective regulatory measures to curb addiction levels from a policymaker's perspective. Employing an infinite-horizon dynamic program that accounts for users' habit-formation behavior, we compare dynamic and fixed pricing across various scenarios. Our findings reveal that dynamic pricing results in higher profits and reduced addiction, benefiting both the firm and users. Similarly, one-time promotions for new users initially foster addiction but ultimately lead to decreased levels in the long run. Furthermore, we analyze addiction control policies such as consumption tax, price lower bounds, and consumption limits. By assessing the profit-addiction efficiency frontier of these regulatory approaches, our study demonstrates that the choice of pricing scheme (fixed or dynamic) and experience randomness significantly influence the optimal regulatory strategy. This research provides valuable insights for tackling digital addiction while preserving profitability.

MC18

Summit - 338

Bundling and Opaque Selling Strategies

Invited Session

Revenue Management and Pricing

Chair: Xiaobo Li, National University of Singapore, Singapore, Singapore

1 - Less is More: a General Framework of Opaque Selling

Jie Liu, School of Data Science, The Chinese University of Hong Kong, Shenzhen, Shen Zhen, China, People's Republic of, Hailun Zhang, Zhenghua Long

Selling opaque products has become popular on e-commerce platforms where the hidden attribute is often color or style, thanks to its obvious advantage in risk pooling the demand. The objective of this study is to quantify the value of the flexibility that underlies opaque selling and investigates the opaque design problem. We first consider a setting in which an online retailer sells a product with N items that only differ in a certain secondary attribute, and in addition, offers opaque options where the customer selects a subset of all products from which the seller allocates one to the customer. We find that the cost savings from using just 2-opaque products is on the same order as the fully flexible case where the opaque product contains all N products, extending the results in Elmachtoub et al. (2019). We next investigate the opaque design problem where the objective is to maximize the expected profit through opaque product design and pricing, and identify conditions under which partial design beats full design. Extensive numerical experiments are conducted to support our main results and insights.

2 - Optimizing Size-Based Bundle Promotions

Ruijiu Mao, National University of Singapore, Singapore, Singapore, Xiaobo Li, Chung Piaw Teo

Current size-based bundle promotions often overlook the diverse needs of customers and variations among products. Recognizing these limitations, we conducted a comprehensive exploration of size-based bundling promotions. We examined key elements such as determining promotion numbers, selecting product pools, setting bundle prices, and estimating customer utility for different products. We studied a choice model for size-based bundle promotions and proposed an efficient polynomial-time assortment planning method under reasonable assumptions. We also compared our results with the heuristics applied by e-tailing companies.

3 - Separation in Distributionally Robust Monopolist Problem

Qiu Hao, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, zhen wang, Simai He

We consider a monopoly pricing problem, where a seller has multiple items to sell to a single buyer, only knowing the distribution of the buyer’s value profile. The seller’s goal is to maximize her expected revenue. In general, this is a difficult problem to solve, even if the distribution is well specified. In this paper, we solve a subclass of this problem when the distribution is assumed to belong to the class of distributions defined by given marginal partial information. Under this model, we show that the optimal strategy for the seller is a randomized posted price mechanism under which the items are sold separately, and the result continues to hold even when the buyer has a budget feasibility constraint. Consequently, under some specific ambiguity sets which include moment-based and Wasserstein ambiguity sets, we provide analytical solutions for these single-item problems. Based on the additive separation property, we show the general additive separation problem is a special case of resource allocation problems that can be solved by known polynomial-time algorithms.

4 - Joint Assortment and Inventory Optimization for the Front-End Warehouse

Kexin Wu, National University of Singapore, Singapore, Singapore, Xiaobo Li, Xue-Ming Yuan

This paper considers the single-period inventory and warehouse assortment planning problem for a two-warehouse system containing a front-end warehouse and a back-end warehouse. In the first stage, before the customers come, the company makes the inventory and assortment decisions for the front-end warehouse. After the customers have placed the orders, the company determines which orders should be fulfilled by the front-end warehouse and which by the back-end warehouse, not considering order splitting as it is cost-prohibitive. Orders assigned to the back-end warehouse incur higher fulfillment costs compared to those fulfilled by the front-end warehouse, primarily due to increased distances and extended waiting times. We formulate the problem as a two-stage stochastic programming model with a total capacity constraint. The primary objective is to minimize the overall product storing fixed cost and fulfillment cost. Optimal fulfillment policies for order structures including pure bundles, series cover bundles, and star cover bundles are proposed. For general order structures, we propose an approximation for the inventory problem and introduce the corresponding approximation algorithm. A tight 1/Z performance bound for the approximation is proved, where the Z indicates the largest number of order types that include a certain product. For two special cases of star cover bundle and sparse demand, we also show tighter performance bounds.

MC19

Summit - 339

Optimization for Portfolio Selection

Contributed Session

Chair: Johannes Norheim, University of Strathclyde, Glasgow, 02139

1 - Prescriptive Bi-level Approach for Portfolio Risk-return Management

Junyao Yang, Governors State University, University Park, IL, United States, Aleksandr Vinel

Bi-level optimization, a class of hierarchical mathematical programming, has gained significant attention in decision-making. In this study, we consider the following broker-investor portfolio problem. An investor is interested in contracting with a broker. Naturally, the broker can select model parameters to adjust his risk preferences and performance-based incentive structure. The investor can be reasonably assured that the broker will maximize the portfolio return. What may be missing is a more concerted effort to minimize risk, due to the mismatch objective between the parties. Therefore, we consider this challenge and provide some modeling and computational insights. We intend to divide the decision space using prediction to introduce a novel view of risk. Specifically, we use machine learning to predict market movement and split the decision space apart into two sub-spaces, bullish and bearish. We assume that the investor will focus only on the bullish set, which formulates an upper-level risk minimization problem. The broker at the lower level is guided by the investor in seeking return, with the constraint of risk ensuring that the lower level risk remains within acceptable bounds. Most importantly, the upper level will have a different decision space from the lower level. The main contributions of this work are as follows: 1) we incorporate a hierarchical bi-level approach with two-level decision-making on portfolio optimization problems 2) we develop a new bi-level optimization formulation that considers modern ML technology in making decisions while performing risk hedging and seeking returns.

2 - Optimizing portfolio performance through Linearized Adaptive Boosting combination

Minsu Chung, Yonsei University, Seoul, Korea, Republic of, Seongmoon Kim

This study introduces a portfolio combination algorithm, named the Linearized Adaptive Boosting combination (LABc), which employs the modified Adaptive Boosting method to combine multiple portfolio models and obtain the optimal portfolio. LABc adjusts its corresponding combination weight according to the Euclidean distance of investment weight for each individual portfolio model, relative to the optimal investment weight determined by the Tangency portfolio. This approach is logical, as a larger Euclidean distance indicates that the individual portfolio model is further from the Tangency portfolio thus less optimal. Through iterative adjustments, LABc dynamically modifies the combination weight by considering both the error term and the importance value, which will be detailed in this paper.

To evaluate the performance of LABc, we conducted an investment experiment using 10 different datasets. The results demonstrate that LABc achieve significantly high Sharpe ratio, Sortino ratio, and a high Calmar ratio across every datasets, compared to 8 different benchmark portfolios. Additionally, LABc maintains a reasonably low turnover ratio, which is crucial for ensuring that the actual profit is not eroded by excessive trading costs. Through extensive experimentation, our study demonstrates that LABc consistently outperforms its benchmark portfolios.

3 - News Sentiment-Based Adjustment to the Markowitz Portfolio Model

Joomi Park, Yonsei University, Seoul, Korea, Republic of, Minsu Chung, Hongseon Kim, Seongmoon Kim

This paper suggests a portfolio selection model that adjusts the Markowitz portfolio selection model based on sentiment analysis results. This model is designed to adjust the weights of investments according to the results of sentiment analysis, making a flexible response to market conditions possible. News headline data were collected from eight major South Korean economic daily newspapers, and sentiment indices were derived using KR-FinBERT. To evaluate the performance of the model, we conducted empirical experiments using historical monthly return data from the South Korean market.

4 - Formulas for Calculating Out-of-Sample Performance of Constrained Portfolios

Luis Chavez-Bedoya, ESAN Graduate School of Business, Lima, Peru

Within the framework of Kan and Zhou (2007), this paper investigates the out-of-sample performance of optimal mean-variance (MV) portfolios under a set of linear weight constraints. These optimal portfolios combine the risk-free asset, the constrained minimum-variance portfolio, and a hedge portfolio with zero exposure to the constraints and maximum performance. Our analysis yields insights that contribute to the development and evaluation of strategies aimed at enhancing out-of-sample performance. Furthermore, our findings offer theoretical results and formulas pertaining to expectations of functions of quadratic forms and random matrices. These results prove valuable for examining out-of-sample returns and performance of constrained MV portfolios.

5 - Portfolio Selection Models with Shrinkage Factors Derived from a Nonlinear Shrinkage Method Reflecting Macroeconomic Variables

Jihye Yang, Yonsei University, Seoul, Korea, Republic of, Seongmoon Kim

We propose portfolio selection models with weight constraints reflecting four important macroeconomic variables: Baltic Dry Index, Interest rate, Money supply, and Inflation rate. We design two types of macro augmented portfolio models (MAPs); MAPs with a linear shrinking factor (MAPsL), MAPs with a nonlinear shrinking factor (MAPsN). Furthermore, we incorporate four nonlinear shrinking factors (NSFs) altogether into the construction of MAP with the combination level derived from a tracking signal (MAPcN). We assess the performance of the MAPs compared to those of benchmarks in 12 different stock markets and find that the MAPs outperformed the benchmarks, with the MAPsN outperforming the MAPsL and the MAPcN outperforming the MAPsN in terms of the Sharpe ratio, Sortino ratio, and Rachev ratio. Furthermore, we evaluate the performance of MAPs against each other. In terms of individual measure as well as collective ones, each of MAPsN outperforms the corresponding constituent of MAPsL. The effectiveness of the nonlinear shrinkage method measured by the ratio of the Sharpe ratio (*RS*) is differentiated by stock market's characteristics. The *RS* for BDI, IR, MS, and INF is closely related to the trade dependence, the openness of financial market, the stability of money supply, and the market concentration, respectively.

6 - Robust Optimization for Technology R&D Portfolios

Johannes Norheim, Massachusetts Institute of Technology, Cambridge, MA, United States, Afreen Siddiqi, Olivier L. de Weck

One of the key challenges in technology roadmapping is the strategic decision-making regarding which technologies to invest in over time. Unlike financial investments, the returns on technology investments are closely tied to the technology's added value in revenue-generating products. However, the lack of extensive historical data on technology complicates accurate performance forecasting, introducing risk with the investment in specific technologies. We introduce a robust optimization methodology for decision-making in technology R&D portfolios to address data sparsity and uncertainty. We demonstrate the effectiveness of our approach with a technology R&D portfolio case study from the aerospace industry. Our results underscore the potential of robust optimization in managing the uncertainties and complexities associated with technology investment decisions, offering a promising avenue for future research and application.

MC20

Summit - 340

Advances in Information Design and Platform Design

Invited Session

Decision Analysis Society

Chair: Qian Zhang, California State Polytechnic University Pomona, Pittsburgh, PA, United States

1 - Rating Systems under Customer Disconfirmation Bias: Convergence and Granularity

Qian Zhang, University of Pittsburgh, Pittsburgh, PA, United States, Arian Aflaki

Customers and platforms increasingly rely on online ratings to gauge the quality of the products and services. Additionally, customers reflect on their own experiences by rating the product. Empirical work has established the presence of reference effects when customers post reviews. In other words, customers' prior beliefs about the product's quality affect their judgment, which is influenced by the previous reviews. In this work, we study the convergence of the ratings to the product's intrinsic quality in the presence of reference effects. In contrast to the case when customers' judgments are absolute, we show that the sequence of ratings affects this convergence with reference effects. Specifically, a rational agent can infer the intrinsic quality from the entire rating history. However, platforms often post only statistics of these ratings, such as their average. In this case, we show that convergence to intrinsic quality is not guaranteed in the presence of reference effects. This creates information asymmetry between the platform and customers.

2 - Inventory Disclosure with Product Returns

Ceren Gultekin, CUNY, Graduate Center, Baruch College, New York, NY, United States, Tolga Aydinliyim

Consumers may strategically defer their purchases in anticipation of future discounts at the risk of a stock-out. Retailers can mitigate the adverse profit implications of consumers' strategic waiting by optimally deciding whether to disclose the inventory availability or mask it by displaying "in stock" messages. We study the optimal inventory disclosure strategy when consumer valuations are uncertain which might trigger returns. We use a price- and refund-setting newsvendor framework with availability-dependent demand. Returned and unsold units from the primary selling period are salvaged during clearance. Under masked inventory, boundedly rational consumers form beliefs about the clearance period availability and make buy now or wait decisions accordingly. We show that contrasting the optimality of consistent disclosure of low availability with known customer valuations and no product returns, selective disclosure can be optimal when uncertain consumer valuations trigger product returns. We characterize the economic conditions that favor the optimality of selective disclosure. Furthermore we prove that selective disclosure is more likely to sustain at equilibrium for high margin products when it complements the retailer's optimal refund decision.

3 - Why Would a Principal Sponsor a Persuasive Agent with Conflicts of Interest

luying wang, Tianjin University, Champaign, China, People's Republic of

In this paper, we examine conflicts of interest that arise in sponsorship scenarios where a principal (he) sponsors a persuasive agent (she) to boost his sales among her followers, particularly when the agent owns a product.
We propose a Bayesian persuasion model allowing the agent to design the structure of the match information between the players' products and followers (i.e., changing the persuader's style or mode of communication).
We find that given payment by the principal, the persuasive agent with product ownership may have a higher Level of Persuasiveness (LP), i.e., the greater extent of the sponsor-orientated manipulation on the information structure, than the one without.
Interestingly, the principal may be more motivated to sponsor an agent with a product than without.
This is particularly the case when there is a strong Direct Persuasive Conflict (DPC) between their respective demands, and an Indirect Persuasive Conflict (IPC) affecting the principal's sales revenue and the agent's information utility.
This result is robust even when the manipulation of the persuasive agent could be self-orientated.
Finally, we find that the sponsorship between the principal and the agent might hurt followers' welfare and social welfare.

4 - Optimal dynamic contracting for teams

Ping Cao, University of Science and Technology of China, Hefei, China, People's Republic of, Peng Sun, Feng Tian, Yangge Xiao

We consider a continuous-time setting of a principal contracting with a team of agents to undertake a project. The success rate of the project depends on the agents' hidden effort level. Assuming that the principal and the agents are all risk-neutral and do not discount the future, we consider three scenarios: cooperation in which the agents work together, noncooperation with team performance in which the agents work separately and the principal can only observe the joint outcome, and noncooperation with individual performance in which the principal can observe each agent's outcome. We derive the optimal contracts for these scenarios, and find that a randomized firing may be required in the latter two scenarios. By comparing the principal's optimal utilities under these scenarios, we obtain some implications for optimal contract design.

5 - The Water Far away Can Put Out the Fire Nearby: Enhancing Healthcare Accessibility through Platform Design

Li Xiang, Peking University, Beijing, China, People's Republic of, Li Ling, Lu Yiping

Telemedicine and telehealth show potential for reducing barriers to access and enhancing quality of healthcare. However, the impact of online healthcare platforms on patients' choices of providers and the resulting spatial distribution of healthcare utilization remains poorly understood. This study assesses the impact of demand shocks—particularly in regions with inadequate healthcare resources—on the flow of online medical consultations using a unique dataset from Baidu Health, a leading online healthcare platform in China. We analyze Baidu search queries to explore demand shocks linked to outbreaks of respiratory illnesses nationwide. Our findings indicate that the platform's design, especially its matching algorithm, significantly contributes to reducing disparities in healthcare resource allocation across regions, as well as between urban and rural areas, and among socioeconomic groups. Furthermore, we utilize an equilibrium model of patient search to quantitatively assess the welfare gains attributed to these matching algorithms, which are found to be 12.0% greater in the presence of demand shocks. These results have important policy implications for enhancing healthcare access in China and potentially other countries.

MC21

Summit - 341

Healthcare Analytics and Decision Support Systems For Improved Patients Outcome

Invited Session

Decision Analysis Society

Chair: Masoud Salehi, Iowa State University, Ames, IA, United States

1 - Identifying Cognitive Decliners using Neuroimaging Data

Samaneh Rezaeimanesh, Oklahoma State University, Stillwater, OK, United States, Mohammad Fili, Parvin Mohammadiarvejeh, Guiping Hu, Auriel Willette

As people age, cognitive decline often becomes a natural aspect of the aging process. However, cognitive trajectories vary among individuals. Identifying Cognitive Decliners is crucial for understanding the factors that protect against or prevent cognitive decline, thereby contributing to successful aging. This study involved 711 participants aged 65 and above from the UK Biobank. Participants were categorized as "Cognitive Decliners" or "Others" based on cognition scores from three tests: Fluid Intelligence (FI), Pairs Matching Memory (PMM), and Reaction Time (RT). We assessed the effectiveness of various Magnetic Resonance Imaging (MRI) modalities, including resting-state functional MRI (rsfMRI), diffusion MRI (dMRI), and structural MRI (sMRI), in identifying Cognitive Decliners.

2 - An Exploratory Analysis on the Effects of Introduction of Electronic Health Record Systems Act on Patients Outcome

Masoud Salehi, Iowa State University, Ames, IA, United States

The adoption of Electronic Health Record (EHR) systems represents a pivotal transition in modern healthcare delivery, promising streamlined data management and improved patient care. However, the implications of the legislative mandates surrounding EHR implementation, such as the Electronic Health Record Systems Act (EHRSA), remain underexplored in terms of their direct impact on patient outcomes. This study embarks on an exploratory journey to elucidate the multifaceted effects of EHRSA implementation on patient health metrics and healthcare quality.

3 - Enhancing Patient Care Through Telemedicine: Linking Primary Care with Specialists

Mohammad Delasay, Stony Brook University, Stony Brook, NY, United States, SALIGRAMA AGNIHOTHRI, Arvind Sainathan

We investigate how using telemedicine and e-consultation platforms to link primary care doctors with specialists can enhance patient care and coordination. We also examine how referral choices and resource allocation are affected by different payment methods.

MC22

Summit - 342

Education Analytics

Invited Session

Decision Analysis Society

Chair: Shreyas Sekar, University of Toronto - Rotman School of Management, Toronto, ON, Canada

Co-Chair: Andre Cire, University of Toronto, Toronto, ON, Canada

1 - The Impact of Race-Blind and Test-Optional Admissions on Racial Diversity and Merit

Allen Sirolly, Columbia Business School, New York, NY, United States, Yash Kanoria, Hongyao Ma

How significant was the role of racial preferences in U.S. college admissions before the Supreme Court’s 2023 decision to ban affirmative action? How much might test-optional admission policies impact racial diversity and academic merit? In this work, we estimate a simple model of college admissions decisions from 2012–2021, leveraging a novel dataset of applicant profiles and admissions outcomes across the full spectrum of college selectivity. We find that, broadly, the impact of race and testing policies on diversity and merit of admits decreases by college selectivity. For America’s less selective colleges that collectively enroll over three quarters of students, fully eliminating racial preferences (expressed either directly or via unobserved correlates) has little impact on the proportion of underrepresented minorities (URM), and on the average SAT score of admitted students. In contrast, for the 34 most selective colleges accounting for 3 percent of total enrollment, our estimates suggest that admissions going "race blind", absent any compensating changes in admissions criteria, could result in a 40–50% decrease in URM admission and a 9–14 point increase in the average SAT score of admits. Similarly, we estimate that universal test-optional admissions may increase URM admissions by 10–15% and decrease the average SAT score by 15–23 points at elite colleges, but may not materially affect the less selective institutions.

2 - A Rolling Recruitment Process Under Applicant Stochastic Departures

Qing Li, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Chenyin Gong

We study rolling recruitment processes in which applicants arrive stochastically over time and are available for a random amount of time. In each period, the recruiter either stops or waits. If they stop, they determine how many offers to make and whom to make offers to, and the applicants who do not receive an offer will leave. The focus of our study is on the effects of applicant qualifications, measured by scores, on the recruiter's optimal policy.

For each applicant's score, there are two thresholds. If the score exceeds the higher threshold, the recruiter stops and makes an offer to the applicant and possibly to others. If the score falls below the lower threshold, the recruiter also stops but makes no offer to the applicant. If the score is in between the two thresholds, the recruiter waits. What is the impact on an applicant's likelihood of receiving an offer if their competitors become more qualified? When the score of another applicant increases, the recruiter may change from making an offer to an applicant to waiting or to instead making an offer to the other applicant whose score has increased. In other words, an applicant may be disadvantaged if they face stronger competitors. However, an applicant may also benefit from having stronger competitors. When the score of another applicant increases, the recruiter may change from not making an offer to an applicant to making an offer to both applicants. We also numerically evaluate the benefit of endogenously determining the stopping rule.

3 - Who Gets to Ride? Comparing Transportation Eligibility Policies for Saanich Public Schools

Olivia Phillips, H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech, Atlanta, GA, United States, Arthur Delarue, Sebastien Martin

In public school transportation, eligibility policies determine which students get to ride the bus and which do not. Determining the operational cost of individual eligibility policies is an intractable problem and school districts must often make decisions with incomplete information. We collaborate with the Saanich School District in British Columbia, Canada, to develop methodologies to evaluate the costs and benefits of various eligibility policies. Our work involves both analyzing the costs and benefits of the current policy as well as designing optimization models that maximize utility while keeping costs below a given threshold.

4 - Beyond the Scores: Incorporating Chance in Sequential Admission Processes

Andre Cire, University of Toronto - Rotman School of Management, Toronto, ON, Canada, Shreyas Sekar

We investigate a sequential decision making problem where a University sends offers to students to achieve a given admission’s target, but only learns the true outcomes of student’s after a number of periods have elapsed. The resulting model is a challenging dynamic combinatorial system due to the uncertain action response and delayed rewards. We propose offline and online policies under a new smoothed penalty function associated with the desired student target, which allow us to derive interpretable ordering and threshold strategies that are suitable to large-scale hiring processes. Our interpretable policies come close to achieving the optimal outcome, and considerably improve on the performance of status-quo greedy policies that are utilized in many organizations. In particular, our work establishes that incorporating candidates' acceptance probabilities into the admissions process can result in significant improvements over existing policies that only prioritize candidates with high scores. We assess the quality of our policies on a large-scale University dataset.

MC23

Summit - 343

Group Decision Making for Sustainable Development

Invited Session

Group Decision and Negotiation

Chair: Ginger Ke, Memorial University of Newfoundland, St. John's, Canada

Co-Chair: Gang Liu, Tianjin University, Tianjin, 300072, China, People's Republic of

1 - A network analysis of external shocks on the dynamics and resilience of the global staple food trade

Huimin Wang, Tianjin University, Tianjin, China, People's Republic of, Guijun Liu

International food trade plays a crucial role in enhancing global food security by connecting regions with diverse agricultural capabilities and resource endowments. This study employs ecological network analysis to investigate the historical dynamics of global staple food trade network resilience from 1986 to 2020. Additionally, structural decomposition and econometric analysis are used to explore the drivers of resilience from both internal and external perspectives. The findings reveal significant heterogeneity in the resilience dynamics of global staple food trade when faced with external shocks, such as COVID-19 and geopolitical tensions. Staple foods such as wheat, rice, and potatoes have demonstrated increased resilience in response to disruptions caused by the pandemic, whereas maize and wheat have been more adversely affected by geopolitical tensions, showing a more pronounced inhibitory effect on their resilience. Potatoes emerge as the most resilient staple, while soybeans exhibit the lowest resilience. Interestingly, the inclusion of a greater variety of staple foods in the aggregated staple food basket does not necessarily give rise to higher resilience. For instance, integrating potatoes, which are characterized by high network efficiency, into the aggregated staple food basket increases resilience, while integrating soybeans with low network efficiency reduces resilience. Furthermore, the diversity of trade flows and trade partners play an important role in enhancing resilience. This comprehensive analysis provides valuable insights for policymakers and stakeholders aiming to bolster the resilience of the global food trade network.

2 - An "Initial-Engineering-Enduse" Quasi-Market Pricing Mechanism under Cascading Risk for Cross-Regional Water Rights Trading

Gang Liu, Tianjin University, Tianjin, China, People's Republic of, Xuxia Li

Water rights trading, operating under a market economy, is recognized as an effective approach to enhance water use efficiency and address water scarcity. It remains a crucial challenge to design a water rights pricing mechanism that incorporates the spatial heterogeneity of water resources. This study aims to address this issue by proposing an “Initial-Engineering-Enduse” quasi-market pricing mechanism based on an empirical study of cross-regional water rights trading in Pingxiang, Jiangxi, China. The proposed mechanism consists of three stages, accounting for associated risks for each stage. Three pricing models, namely stochastic differential game, incomplete contract theory, and the Nash-bargaining model, are calibrated for the three stages based on our study. The interplay among the three equilibrium water prices in each stage is examined to understand cascade effects. It is worth noting that the local government of Pingxiang has adopted and implemented this pricing mechanism. Empirical results demonstrate the following key findings: (1) The quasi-market pricing mechanism effectively reveals the determination of enduse water prices. (2) The government's initial investment can mitigate trading risks and enhance the stability of water rights trading. (3) Neglecting the value of ecosystem services results in an undervaluation of both initial water rights prices and enduse water prices. Furthermore, the government should consider making initial investments in water conservancy projects to minimize risks. Overall, this study provides a practical and promising approach to pricing cross-regional water rights trading in both water-rich and water scarce areas.

3 - Hierarchical determinants of farmer’s water conservation decisions with bounded rationality

Yi Xiao, Tianjin University, Tianjin, China, People's Republic of, Xinran Guo

Despite significant advancements in water conservation technologies, their underutilization by farmers poses a challenge in achieving deeper conservation. Farmers’ interests and decision-making processes are crucial for the successful and effective implementation of water-saving policies considering their essential roles in agricultural production, irrigation, and making water-saving decisions. This research focuses on the micro-perspective of individual farmers, aiming to understand the influence path of various determinants on farmers' water-saving decisions and simulate the impact of incentive policies based on these influence pathways. Using PRISMA statement and meta-analyses, key factors influencing farmers' water conservation decisions are identified from diverse sources including journal articles and government reports. The hierarchical structure of these factors is outlined using the fuzzy interpretive structural model. Then a water conservation decision framework is constructed to characterize the decision-making mechanism of farmers with bounded rationality, which consisting of a decision model for individual farmers, and a dynamic evolutionary game model among multiple farmers. The interactions between the societal and hydrological systems are simulated based on the above models, so that the influence of various water-saving policies on the overall adoption rate of water-saving measures and ecological effects can be investigated. Key sensitive determinants are identified and the pathways of their impact on water-saving behavior could be understood. Recommendations for intervention pathways through new and existing policies and guiding strategies can be proposed. The findings provide valuable insights for enhancing water conservation efforts within agricultural practices, and supporting sustainable water management in critical regions.

4 - Research on the coordination mechanism of platform ecosystem value co-destruction conflict from the perspective of stakeholders

Yu Han, School of Business, Changzhou University, Changzhou, China, People's Republic of

Based on the theory of graph model for conflict resolution, this project (1) investigates and judges deviation. Consider the value preferences of stakeholders, determine the ideal goal of multi-benefit win-win platform value co-creation development, and analyze the deviation between the platform operating status and the ideal situation; (2) designs coordination mechanism. According to the deviation, a multi-objective conflict coordination integer programming model of platform system is constructed. The objective function of the model is determined according to the minimization of the deviation between the preference that can achieve the ideal situation and the current preference. It tries to solve the value conflict problem under the joint action of human stakeholders and non-human stakeholders.

The research results are intended to provide theoretical reference and decision-making suggestions for platform enterprise efficiency, consumer improvement, government and other functional departments to enhance management effectiveness and stabilize platform market operation order.

MC24

Summit - 344

Advanced Analytics: Insights and Innovations

Contributed Session

Chair: Ezgi Eren, PROS, Inc., Houston, TX, United States

1 - Bite by Byte: An open-source automatic bite detection system from mealtime videos in children

Yashaswini Rajendra Bhat, Pennsylvania State University, State College, PA, United States, Timothy Brick, Kathleen Keller, Alaina Pearce

Our innovative, open-source bite detection system advances eating behavior research by utilizing deep neural networks. This system can overcome the challenges associated with traditional manual annotations of bite, which are error-prone, time-consuming, and expensive, thus limiting their practicality for large-scale studies. Unlike available automatic sensor-based tools such as acoustic and oral strain sensors, alter naturalistic eating behaviors and have low accuracy, our system can be utilized in preserving natural eating behaviors.

We employ adapted Faster R-CNN for face detection (80% accuracy) and utilize AlexNet and LSTM models for bite classification. Our dataset consists in-lab recorded meal videos of children (n=345 meal videos), split into a 70:15:15 split for training, validation, and testing.

This approach enhances bite detection in small-scale labs, facilitating quicker annotations for eating behavior studies. It advances the field by providing a scalable, user-friendly tool for health outcome analysis and personalized, adaptive interventions based on live eating behavior.

2 - Which machine learning models resemble human perception? An empirical analysis through the uncanny valley

Woojae Kim, Yonsei university, Seoul, Korea, Republic of, Kyuho Maeng, Youngsang Cho

Many researchers are currently working on improving the accuracy of image classification. However, in the perspective of human robot interaction, especially in uncanny valley, whether artificial intelligence (AI) can measure the similarity of robot's similarity like human is important because similarity recognized by human is used as criterion of uncanniness. Therefore, we conducted the experiment to investigate which feature extraction models have human's perspective. To reflect our hypothesis that the distance of robot image from the hyperplane of SVM will increase as the robot image is not similar to human, we used SVM based on distance-based similarity concept. For data, we used the human image data from Flickr-Faces-HQ Dataset (FFHQ) and robot image data from A-bot database and Mathur and Reichling's study. For feature extraction models, we used CNN, CNN-PCA and derived the distance by the SVM with extracted features. Then, we compared the robot images' similarities that previously responded by people in the survey with the distance from the hyperplane of the SVM. Through the experiment, it is expected that effective technology development will be possible to determine whether the robot appearance in the development stage will be located in an uncanny valley.

3 - A weighting local search for extended binary integer programs

Shunji Umetani, Recruit Co., Ltd., Tokyo, Japan

We present a weighting local search algorithm for a class of binary integer programs including a wide variety of assignment problems. The quality of locally optimal solutions typically improves using a well designed neighborhood search specified for the problem to be solved. However, we have few alternatives other than k-filp neighborhood search due to few useful features for improving the efficiency of algorithms from the general form of binary integer programs. To overcome this, we consider an extended formulation of the binary integer programs that introduces additional data for deriving a specially designed neighborhood search for a class of specific combinatorial structures. We introduce data of bipartite graph in addition to the binary integer programs that derives shift and swap neighborhood search used in a wide variety of assignment problems. The extended formulation enables us to solve a wide variety of assignment problems with both high performance and versatility; e.g., we can easily introduce additional side constraints into the generalized assignment problem with high performance and it makes great advantages to tackle real industrial applications.
We report our computational results of the proposed model and algorithm for a number of test instances.

4 - Data Engineering and Machine Learning Made Easy with KNIME Analytics Platform

Dursun Delen, Oklahoma State University, Tulsa, OK, United States

In the rapidly evolving field of data science, the ability to simplify complex tasks is paramount. In this presentation, you will learn about the KNIME Analytics Platform, a free, open source, no-code tool that makes advanced data engineering and machine learning accessible to everyone. By eliminating the need for extensive coding knowledge, KNIME empowers a broader audience (i.e., citizen data scientists) to engage with intricate processes such as data wrangling, data blending, and machine learning.

5 - Adaptive Data-Driven Revenue Management and Integration with Pricing

Ezgi Eren, PROS, Inc., Houston, TX, United States, Jiabing Li

Traditional revenue management relies on long and stable historical data and predictable demand patterns, which are not possible to meet for certain industries with high demand volatility or relatively shorter history of revenue management practice. Even for passenger airlines where revenue management is the most established, susceptibility to external shocks is a well-known challenge, which often necessitates user monitoring and intervention. We present a robust data-driven approach to revenue management which eliminates the need for demand forecasting and mathematical optimization techniques. We develop a methodology to generate bid prices using only historical booking data without the need for historical controls. Our approach is an ex-post greedy heuristic to estimate proxies for bid prices as a function of remaining capacity and time to departure solely based on historical booking data. We utilize a neural network algorithm to project bid price estimations into the future. An extensive simulation study shows that our methodology is robust to demand shocks as the revenue stays within 1% of the theoretical optimum for a wide range of deviation of expected demand from true demand parameters. We also present results from simulations where we highlight the strategy used for integrating revenue management into pricing is critical for optimal revenue performance.

MC25

Summit - 345

Gams/D-Wave

Invited Session

Technology Showcase

1 - Optimization pipeline design: from data curation to algebraic modeling with GAMSPy

Adam Christensen, GAMS Development Corp, Fairfax, VA, United States, Atharv Bhosekar

Algebraic modeling languages (AMLs) have been a cornerstone in the fields of optimization and economics. These tools are popular because they are able to effortlessly link the worlds of algebra and computer science — that is, the syntax of the AML closely approximates that of handwritten algebra but its execution is automated and scalable. AMLs, by design, are not general purpose programming languages; as a result, it can be difficult to gather, clean and prepare data for a modeling environment. Recent years have seen sophisticated data science tools enter the mainstream. Languages such as Python and R can leverage Numpy/Pandas and Shiny/Tidyverse/Dplyr to efficiently work with large data in deployable environments. Modern infrastructure tools such as Docker and Kubernetes make it possible to isolate workflows and scale compute resources via cloud platforms. All of these compute resources mean that data assets are arriving at optimization model instances from an ever-diversifying number of start points. In this workshop, we present a Python package called GAMSPy that leverages modern data science tools with the flexible nature of Python to construct a powerful Python-AML. This presentation will cover a number of real-world inspired examples that illustrate how to bring data into an environment and effectively clean it for use in an optimization model.

2 - Quantum Optimization: Applying Quantum Computing to Business Optimization Challenges

Murray Thom, D-Wave, Vancouver, BC, Canada

A recent Hyperion Research survey reveals that businesses expect up to 20x return on investment from quantum optimization investments, and over 21% plan production-level use of quantum computing over the next 12-18 months–a 50% increase over the past two years.

Attend this session to learn about quantum optimization, and how these new resources are being used for challenging combinatorial optimization problems such as:
• Optimized workforce scheduling for improved employee experience
• Enhanced production scheduling to improve customer satisfaction
• More efficient and more sustainable logistics routing.

This content is most relevant to Associate (Early Career), Professional (Mid-Career), and Executive (Senior Level).

MC26

Summit - 346

Advanced Algorithms and Stochastic Modeling in Complex Systems

Contributed Session

Chair: Elif Rana Yoner, Bilkent University, Ankara, 06800, Turkey

1 - Enhanced Quantification of Renewable Carbon in Co-Processed Fuels Using Interpretable Segmentation Modeling

Alimzhan Kussainov, University of South Florida, Tampa, FL, United States, Liang Cao, Kaixun Hua

To address the challenge of accurately quantifying renewable carbon content in co-processed fuels, we propose a method that employs an optimal decision tree model. Traditional approaches apply feature selection uniformly across the entire dataset, failing to adapt to different operational conditions. Our method enhances accuracy by partitioning the dataset into subsets guided by the decision tree model. Each subset then independently develops a robust Bayesian linear regression model, utilizing features identified along the specific decision path. This tailored approach has proven effective in real-world oil refining datasets, providing a precise tool for assessing renewable fuel content.

2 - Using Characteristic Functions of Binary LÉVy Processes for Degradation-Based Reliability Analysis

Yu Shi, Tarleton State University, Stephenville , TX, United States, Qianmei Feng

Engineering systems frequently undergo complicated degradation processes compounded by random shocks, introducing risks to their operational reliability. The analysis of inter-dependence among multiple degradation processes remains a challenging in the field of reliability engineering. One effective approach to model the inter-dependency and the stochastic jumps among degradation processes is by using multi-dimensional Lévy processes. However, a significant hurdle arises in determining the closed mathematical form of the reliability function under Lévy measures and distributions. This research addresses this challenge by investigating the Lévy measure and the characteristic function of multiple dependent degradation processes. The dependence across all dimensions is described by a Lévy copula and the associated multi-dimensional Lévy measure. For a two-dimensional case, we derive the reliability function and probability density function from the characteristic function and the inverse Fourier transform. Numerical examples are provided to illustrate the applications of proposed models for reliability and lifetime analysis of multi-dimensional degradation processes in engineering systems.

3 - Carousel greedy algorithms for the minimum stretch spanning tree problem

Bruce Golden, University of Maryland-College Park, College Park, MD, United States, Jiaqi Wang, Carmine Cerrone

The minimum stretch spanning tree problem aims to find a spanning tree that minimizes the maximum ratio of the distance in the spanning tree to that in the original graph between each possible pair of vertices. Existing heuristic algorithms for this problem are slow and often result in large optimality gaps. To tackle this challenge, we introduce a straightforward and promising carousel greedy algorithm. By investigating the properties of the problem, we further enhance the algorithm's performance. The enhanced algorithm significantly outperforms the best-known algorithms in the literature in terms of both solution quality and running time.

4 - A Decomposable Branch-and-Price Formulation for Optimal Classification Trees

Elif Rana Yoner, Bilkent University, Ankara, Turkey, Ozlem Karsu, Taghi Khaniyev

Construction of Optimal Classification Trees (OCTs) via modeling them as mixed-integer programs which can be solved to optimality, is possible, yet computationally quite cumbersome, especially as the size of the dataset and the maximum depth worked with grows. Our research presents an approach to solving these involved programs to optimality using a novel Branch-and-Price algorithm. We model the classification tree using a "pattern-based" formulation, deciding which feature each leaf must be split upon at each level's decision node. Our preliminary results showcase promising runtimes, displaying the power of decomposition in the realm of binary OCTs.

5 - Epidemic spreading in group-structured populations

Filippo Radicchi, Indiana University, Bloomington, IN, United States

Individuals involved in common group activities/settings – e.g., college students that are enrolled in the same class and/or live in the same dorm – are exposed to recurrent contacts of physical proximity. These contacts are known to mediate the spread of an infectious disease, however, it is not obvious how the properties of the spreading process are determined by the structure of and the interrelation among the group settings that are at the root of those recurrent interactions. Here, we show that reshaping the organization of groups within a population can be used as an effective strategy to decrease the severity of an epidemic. Specifically, we show that when group structures are sufficiently correlated – e.g., the likelihood for two students living in the same dorm to attend the same class is sufficiently high – outbreaks are longer but milder than for uncorrelated group structures. Also, we show that the effectiveness of interventions for disease containment increases as the correlation among group structures increases. We demonstrate the practical relevance of our findings by taking advantage of data about housing and attendance of students at the Indiana University campus in Bloomington. By appropriately optimizing the assignment of students to dorms based on their enrollment, we are able to observe a two- to five-fold reduction in the severity of simulated epidemic processes.

6 - Controlling Tail Risk in Online Ski-Rental

Thomas Lavastida, University of Texas at Dallas, Richardson, TX, United States, Michael Dinitz, Sungjin Im, Benjamin Moseley, Sergei Vassilvitskii

The classical ski-rental problem admits a textbook 2-competitive deterministic algorithm, and a simple randomized algorithm that is $e/(e\text{-}1)$-competitive in expectation. The randomized algorithm, while optimal in expectation, has a large variance in its performance: it has more than a 37% chance of competitive ratio exceeding 2, and the change of the competitive ratio exceeding n is $\Theta(1/n)$!

We ask what happens to the optimal solution if we insist that the *tail risk,* i.e., the chance of the competitive ratio exceeding a specific value, is bounded by some constant δ. We find that this additional modification significantly changes the structure of the optimal solution. The probability of purchasing skis on a given day becomes non-monotone, discontinuous, and arbitrarily large (for sufficiently small tail risk δ and large purchase cost *n*).

MC27

Summit - 347

TIMES Best Working Paper Award 2024

Award Session

Technology, Innovation Management and Entrepreneurship

Chair: Sina Khorasani, Vanderbilt University, Franklin, TN, United States

Co-Chair: Zhaohui Jiang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Exploit or Explore? An Empirical Study of Resource Allocation in Research Labs

Ran Zhuo, University of Michigan, Ann Arbor, MI, United States

Balancing exploitation and exploration in resource allocation under incomplete information is a classic problem in operations management theory. Yet little research has empirically studied how and how well decision-makers make the exploitation-exploration tradeoff in a complex real-world situation. This paper empirically studies how a group of large publicly funded research labs traded off the exploitation of safe projects to maximize short-term productivity versus the exploration of high-variance projects to acquire information and improve long-term productivity. Using granular data on the allocation of almost one million input bundles to more than 300,000 research projects from 2000 to 2015, we model the resource allocation process as a multi-armed bandit and estimate a dynamic structural model to reveal how these labs balanced exploitation and exploration. We find the labs' decision model strongly resembles a simple Upper Confidence Bound (UCB) index. Estimates of the model's free parameters suggest that the labs explored extensively. Counterfactual simulations show that exploration substantially increased the labs' productivity—had they not explored, their output quantity would have decreased by 51%, and their citations would have decreased by 57%. Further simulations demonstrate that the labs' decision model outperformed popular alternative allocation models, including the Gittins Index, Thompson Sampling, and Explore-Then-Commit. Addition- ally, processes that promoted information utilization during allocation contributed to better outcomes. Had the labs not collected and performed data analytics on the information revealed during exploration, they would have saved 3% of funding but lowered output quantity by 7% and citations by 9%.

2 - Using Artificial Intelligence for Reducing Food Waste in Commercial Kitchens

Yu Nu, Cornell University, New York City, NY, United States, Elena Belavina, Karen Girotra

In this study, we estimate the reduction in food waste from the deployment of a system that digitally records instances of food items discarded in a commercial kitchen. We also shed light on the mechanisms that drive this impact. In a quasi-experimental setting, where the system was deployed in ~900 kitchens in a staggered manner, we estimate the impact using synthetic difference-in-differences method. We find that three months after adoption, kitchens generate 29% lower food waste, on average, than they would have in absence of the system— without any corresponding reductions in sales. Utilizing a long-short-term- memory fully-convolutional-networks classifier, we document that these reductions are accompanied by a 23% decrease in demand chasing, a known bias in human inventory management. Upgrading to a system that uses computer vision for waste classification leads to a further 30% reduction in food waste generated by the kitchen, a year after the upgrade. This further reduction is due to accurate recording of infrequent but very high-impact instances of food wasted that are often not entered in the the manual classification system. We also observe substantial effect heterogeneity. Smaller kitchens and those with buffet service (vs. table service) experience almost double the reduction in food waste from the adoption of the system and also from the computer vision upgrade. Low and high-demand-variability sites have higher reductions from adoption than those with medium-demand-variability (42% vs 25%). The impacts of the upgrade are not detectably different with different demand variability.

3 - Optimal Prototyping on Experimentation Platforms

Sanjiv Erat, University of California-San Diego, La Jolla, CA, United States, Sreekumar Bhaskaran

Testing and prototyping comprise an integral part of almost any new product development process. Recent emergence of experimentation platforms who specialize in offering evaluation/assessment ofprototypes to outside parties have opened up the possibilities for firms in reconfiguring their product development processes. Firms can, by outsourcing the evaluation stage of their dev-test cycles, obtain cost efficiencies and scale. At the same time, lack of visibility into the actual evaluation process andthe possible ambiguity in the product development firm's requirements may lead to potentially noisy evaluations, raising concerns regarding the fidelity and accuracy of results. The current article formulates a model of "outsourced-evaluations" and examines howthe noisy low-fidelity evaluations alter the firm's optimal prototyping. Our results demonstrate imperfect evaluation fidelity changes the client's optimal experimentation, with starkest difference being that it can make optimal for the client to select and launch a prototype that did not yield the best evaluation. In addition, our analysis reveals the client should optimally request most precise measurements when their exante uncertainty is moderate. Finally, examining the optimal measurement technology choices ofthe platform, we find that when the client's exante uncertainty increases from moderate to high values, the platform should offer lower fidelity evaluations, butat a higher fee. We develop managerial insights for the optimal choice of fidelity and the optimal length of the evaluation cycle should be planned depending on the platform's evaluation fees and the client's exante uncertainty. The resulting framework can offer guidance to product and software development firms who leverage external experimentation platforms.

4 - Managing Multihoming Workers in the Gig Economy

Park Sinchaisri, University of California Berkeley, Berkeley, CA, United States, Gad Allon, Maxime Cohen, Kenneth Moon

Gig economy platforms compete to source labor from common pools of workers, who multihome by dynami- cally allocating their services in real-time across multiple platforms. The question of how such workers choose between competing platforms has grown in salience. However, the unavailability of comprehensive data has limited our understanding of workers' dynamic multihoming decisions and their impact on the labor supply and operations of platforms. We address this gap by integrating a major ride-hailing platform's proprietary data regarding individual drivers' detailed trips with public data on the drivers' outside options. Using empirical methods that overcome the remaining data limitations using simulation and machine learning, we structurally estimate the perceived costs that motivate drivers' forward-looking decisions and behavior. Our analysis reveals that workers are short-sighted and value sequences of consistent rewards (e.g., jobs and pay) over highly variable ones. Based on these findings, we explore the impact of compensation and incentives on the platform's labor supply and operations. Our counterfactual analyses indicate that consis- tent pay dominates variable pay in retaining multihoming workers. For gig platforms seeking to maintain a stable platform workforce, they can further control their labor supply by rewarding uninterrupted work or delaying quits. For policymakers, our research gives insights regarding the design of gig economy incentive schemes and regulations, including specifically New York City's 2018 Driver Income Rules, and their impact on multihoming behavior.

MC28

Summit - 348

Strategies and Insights on Workforce Training

Contributed Session

Chair: Aqib Siddiqui, IE Business School, Madrid, Spain

1 - Combating Textual Toxicity in MOBA Games

Sunan Qian, University of Notre Dame, South Bend, IN, United States, Xinxue (Shawn) Qu, Corey Angst, Nicholas Berente

Multiplayer Online Battle Arena (MOBA) games such as League of Legends (LoL) and Defense of the Ancients (DotA) attract millions of users globally. Online games provide rich environments for social interaction, where players can collaborate, compete, and establish relationships. However, the emergence of toxic behavior among players has become a significant concern in online gaming, especially for MOBAs that are more competitive. Generally, two aspects of toxic behavior can manifest in the gaming environment – abusive communications (e.g., harassment, verbal abuse, and flaming) and disruptive gameplay (e.g., griefing, spamming, and cheating), all of which fall under the broad notion of toxicity. Such toxic behavior can degrade the gaming experience, impact players retention, and even harm engagement in the gaming community. Understanding and mitigating toxic behavior is crucial for maintaining a positive gaming environment. In this study, we mainly focus on the textual toxicity in chat messages in MOBA games using a rich dataset collected from Dota 2. First, we aim to understand factors that may drive textual toxicity in MOBA games. Second, we plan to investigate both short-term and long-term impacts on players' gaming behavior and community engagement. Lastly, we intend to explore potential interventions that platforms can adopt to mitigate toxic behavior and promote positive playing.

2 - Growing Up Digital: Decoding Age and Gender Influences on Digital Maturity with Explainable AI

Aqib Siddiqui, IE Business School, Madrid, Spain, Konstantina Valogianni, Nancy Pouloudi

In today's digital landscape, fostering healthy Information and Communication Technology (ICT) usage among youth has emerged as a pressing necessity. Digital maturity, a multidimensional construct encompassing a range of competencies, enables young individuals to navigate the digital realm effectively and responsibly. This study investigates the interplay among age, gender, and various dimensions of digital maturity, while also exploring the potential influence of socio-economic status on these competencies. Utilizing a comprehensive survey instrument, we employ Explainable Artificial Intelligence (XAI) technique, SHapley Additive exPlanations (SHAP), to quantify the impact of each digital maturity dimension on overall scores. Our analysis unveils distinct age and gender patterns in dimension salience. For instance, younger children (11-14) prioritize Respect Towards Others, Risk Awareness, and Digital Literacy, while older adolescents (15-18) value Individual Growth, Digital Citizenship, and Support Seeking Behavior. Gender stereotypes also influence dimension importance; Digital Literacy and Individual Growth are consistent across genders, but Regulation of Aggressive Impulses and Support Seeking Behavior vary. We emphasize the need to investigate the socio-cultural contexts shaping digital maturity, aiming to further explore socio-economic status's role. These findings carry implications for educational policies, technology design, and social interventions, advocating for tailored strategies that acknowledge diverse developmental trajectories and socio-economic influences. This study contributes to Information Systems research by offering empirical evidence of dynamic digital competencies, advocating for a holistic approach to technology use. Ultimately, our research advocates for an inclusive digital environment ensuring all youth, regardless of demographics, can thrive in the digital age.

MC29

Summit - 420

Recent Advances in Mixed-Integer Nonlinear Programming

Invited Session

OPT: Integer and Discrete Optimization

Chair: Zedong Peng, Massachusetts Institute of Technology, Cambridge, MA

1 - Recent Developments in the Gurobi Global Minlp Solver

Gregory Glockner, Gurobi Optimization, Bellevue, WA, United States, Edward Rothberg

This talk will focus advances in the Gurobi Global MINLP Solver, both regarding solver performance and model creation. Modeling improvements primarily involve the addition of direct support for modeling nonconvex multivariate nonlinear functions. Performance improvements include better numerical stability, additional presolve reductions, and faster times to the first feasible solution on a significant number of models.

2 - Exploiting Structure in Hybrid Control Problem by Linking Consecutive Periods

Jisun Lee, UC Berkeley, Berkeley, CA, United States, Andres Gomez, Alper Atamturk

We study a mixed-integer quadratic programming formulation of an n-period hybrid control problem with a convex quadratic cost function and a linear system. For the one-dimensional case, we propose a dynamic programming algorithm that solves this problem as a shortest path problem on a directed acyclic graph in polynomial time. Subsequently, we derive an equivalent SOCP formulation in an extended space using path constraints. We discuss how to generalize these results to multi-dimensional cases with potentially side constraints. Finally, we present computational results testing the proposed formulations.

3 - A Spectral Outer Approximation Algorithm for Binary Semidefinite Problems

Daniel de Roux, Carnegie Mellon University, Pittsburgh, PA, United States, David Bernal

This work introduces a spectral second-order outer approximation algorithm to solve to optimality integer semidefinite programs that are exact formulations of binary quadratically constrained quadratic problems. Our approach is inspired by polyhedral and second-order representable regions that outer approximate the feasible set of a semidefinite program relying on a spectral decomposition of a matrix that simultaneously diagonalizes the objective matrix and an aggregation of the constraint matrices. We give substancial evidence that our approach solves larger instances of binary semidefinite problems than the available software alternatives.

4 - Hybrid Quantum Branch-and-Bound Method for Quadratic Unconstrained Binary Optimization

David Bernal Neira, Purdue University, USRA/NASA, West Lafayette, IN, United States, Zedong Peng, Pedro Maciel Xavier

This work presents an enhancement of branch-and-bound (B&B) methods for Quadratic Unconstrained Binary Optimization (QUBO) problems using quantum and classical heuristics as a source of primal solutions. The enhancements in performance if commercial B&B codes when incorporating these heuristics is quantified in reduction of explored notes and wall clock time. Careful consideration is taken when deciding when the heuristics are to be applied, in particular incorporating the limitations of these heuristics into the call-back procedure.

Experiments carried with the commercial B&B solver Gurobi enhanced with classical heuristics such as those implemented in the MQLib and quantum methods such as the D-Wave quantum annealer provide the avenue of accelerating the performance of Gurobi noticeably when aiming to solve QUBO problems to optimality. The implementation is done through the QUBO.jl software package, efficiently integrating Ising/QUBO solvers with mathematical programming solvers available through JuMP.

5 - Automating Idealness Proofs for Binary Programs with Application to Rectangle Packing

Jamie Fravel, Virginia Tech, Blacksburg, VA, United States, Robert Hildebrand

We develop an optimization framework for identifying ideal Mixed Binary Linear Programs (MBLP) which is linear when using known input data and nonconvex quadratic over parametric input data. These techniques are applied to various formulations for rectangle packing, conjectured to be pairwise-ideal. Additionally, we address a variation of the rectangle packing problem which incorporates clearances along selected edges of the packed objects. We present both existing and novel MBLP formulations for the underlying disjunctive program and investigate the poor performance of Gurobi's default branch-and-cut methodology. We operate under a strip-packing objective that aims to minimize the overall height of the packed objects.

MC30

Summit - 421

Optimization Software Session I

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelmann, Arizona State University, Tempe, AZ, United States

1 - Recent Advances in the Cardinal Optimizer

Gerald Gamrath, COPT GmbH, Berlin, Germany

In this talk, we present the recent developments in the Cardinal Optimizer (COPT). We discuss some key techniques that contributed to the performance improvements of our MIP solver and present performance numbers of the latest COPT release for all problem classes.

2 - Advancements in Mindopt: Enhancements in Mixed Integer and Nonlinear Solver

Kuo-Ling Huang, Alibaba, Bellevue, WA, United States

We will present the latest enhancements made to MindOpt, with a specific emphasis on our advancements in Mixed Integer Programming (MIP) and nonlinear solvers. We will discuss the innovative techniques and algorithmic strategies that have been integrated to improve both performance and reliability.

3 - Recent Developments in Optverse Optimization Solver

Zirui Zhou, Huawei Technologies Canada, Burnaby, BC, Canada

In this talk, we present latest development of Optverse optimization solver and some of its successful applications. We will also discuss some new techniques in mixed-integer linear programming and infeasibility detection that improve the performance of Optverse optimization solver.

4 - Recent Developments in the Gurobi Optimizer

Edward Rothberg, Gurobi Optimization, LLC, Beaverton, WA, United States, Ed Klotz, Gregory Glockner, Xavier Nodet

This talk will discuss new features and enhancements in Gurobi 12, including performance improvements. While considering all problem types that Gurobi solves, the talk will focus on the advances in the Gurobi Global MINLP Solver, including its direct support for modeling nonconvex multivariate nonlinear functions.

5 - LeOPT: A Smart Optimizer

wenli ouyang, Lenovo, Beijing, China, People's Republic of

LeOPT is a smart optimizer developed by Lenovo Research AI Lab, currently excels in solving general linear programming and mixed integer linear programming problems. Its modular design and development encompasses various components such as the simplex method, interior point method, branch and bound, cut separation, conflict analysis, and primal heuristics. Crucially, this modular, cross-language implementation enables seamless integration with machine learning models. With the progress in combinatorial optimization driven by artificial intelligence, LeOPT empowers smarter and faster problem-solving.

MC31

Summit - 422

Recent Advances in Statistical Learning and Optimization

Invited Session

OPT: Machine Learning

Chair: Moise Blanchard, MIT, Cambridge, MA, United States

1 - Online Convex Optimization with Prediction Through Accelerated Gradient Descent

Junhui Zhang, MIT, Cambridge, MA, United States, Patrick Jaillet

We study online convex optimization with predictions -- at each time step t, predictions about the next k steps are available -- and coupled costs over time step -- the cost function at time step t depends on the decisions made between time $t-a$ and $t+b$ for some nonnegative integers a,b.

We provide a general recipe to run synchronous update in an asynchronous fashion that respects the sequential revelation of information. Combined with existing convergence results for convex optimization using inexact first-order oracle, we show that acceleration is possible in this framework, where the dynamic regret can be reduced by a factor of $(1-O(\sqrt{\kappa}))^{\frac{k}{a+b}}$ through accelerated gradient descent, at a cost of an additive error term that depends on the prediction accuracy. This generalizes and improves the $(1-\kappa/4)^k$ factor obtained by Li and Li (2020) for $a+b = 1$. Our algorithm also has smaller dependency on longer-term prediction error.

2 - Extracting Trustworthy Rule Ensembles for Interpretable ML

Brian Liu, MIT, Cambridge, MA, United States, Rahul Mazumder

We present an optimization framework to extract sparse representative subsets of decision rules from tree ensembles, that are easy for a practitioner to examine. These frameworks have new functionalities such as penalties to encourage fusion and stability in the selected rules. Optimization problems in our framework are challenging to solve due to problem scale and the non-convexity of our regularization penalties. As such, we develop specialized optimization solvers to tackle problems in our framework, that can solve problems beyond the capabilities of off-the-shelf optimization software. We show in our experiments that rule ensembles extracted using our framework can outperform existing algorithms for constructing sparse rule sets in terms of both predictive performance and interpretability.

3 - Benign overfitting and adaptive nonparametric regression

Julien Chhor, Toulouse School of Economics, Toulouse, France

Benign overfitting is a counter-intuitive phenomenon recently discovered in the context of deep learning. It has been experimentally observed that in certain cases, deep neural networks can perfectly overfit noisy training data, while still achieving excellent generalization performance for predicting new data points. This goes against the conventional statistical viewpoint which posits that there should be a necessary tradeoff between bias and variance. This paper aims to understand benign overfitting in the simplified setting of nonparametric regression. We propose using local polynomials to construct an estimator of the regression function with the following two properties. First, this estimator is minimax-optimal over Hölder classes. Second, it is a continuous function that interpolates the set of observations with high probability. The key element of the construction is the use of singular kernels. Moreover, we demonstrate that adaptation to unknown smoothness is compatible with benign overfitting: indeed, we propose another interpolating estimator that achieves minimax optimality adaptively to the unknown Hölder smoothness. Our results highlight that in the nonparametric regression model, interpolation can be fundamentally decoupled from the bias-variance tradeoff.

4 - On the Robustness of Spectral Algorithms for Semirandom Stochastic Block Models

Naren Manoj, Toyota Technological Institute of Chicago, Chicago, IL, United States

In a graph bisection problem, we are given a graph G with two equally-sized unlabeled communities, and the goal is to recover the vertices in these communities. A popular heuristic, known as spectral clustering, is to output an estimated community assignment based on the eigenvector corresponding to the second-smallest eigenvalue of the Laplacian of G. Spectral algorithms can be shown to provably recover the cluster structure for graphs generated from probabilistic models, such as the Stochastic Block Model (SBM). However, spectral clustering is known to be non-robust to model mis-specification. Techniques based on semidefinite programming have been shown to be more robust, but they incur significant computational overheads. We study the robustness of spectral algorithms against semirandom adversaries. Informally, a semirandom adversary is allowed to "helpfully" change the specification of the model so that's consistent with the ground-truth solution. Our semirandom adversaries in particular are allowed to add edges inside clusters or increase the probability that an edge appears inside a cluster. Semirandom adversaries are useful tool to determine the extent which an algorithm has overfit statistical assumptions on the input. On the positive side, we identify a wide range of semirandom adversaries under which spectral bisection using the unnormalized Laplacian is strongly consistent, i.e., it exactly recovers the planted partitioning. On the negative side, we show that in many of these settings, normalized spectral bisection outputs a partitioning that makes a classification mistake on a constant fraction of the vertices. FWe demonstrate numerical experiments that complement our theoretical findings.

MC32

Summit - 423

Theory and Applications of Stochastic Programs

Invited Session

OPT: Optimization Under Uncertainty

Chair: Nan Jiang, Georgia Tech, Atlanta, GA, United States

1 - Range Value-at-Risk under Distributional Ambiguity: Tight Bounds with Support Information

Zhi Chen, The Chinese University of Hong Kong, Shatin, Hong Kong, Zhenyu Hu, Guangwu Liu, Ruiqin Wang

Estimating risk measures, such as value-at-risk (VaR) and conditional value-at-risk (CVaR), plays an important role when making decisions with limited information. In this paper, we examine the problem of bounding a more general class of risk measures called, range value-at-risk (RVaR), over mean-variance and Wasserstein ambiguity sets. While existing results all assume unbounded support, we characterize tight upper and lower bounds when there is support information. Our numerical results further demonstrate the significant value of support information.

2 - Chance-Constrained Maximum Clique Problem

Shunyu Yao, University of Arizona, Tucson, AZ, United States, Neng Fan, Pavlo Krokhmal

Finding sufficient conditions for the existence of a subgraph H within a larger graph G has a rish story in graph theory and discrete mathematics. Random graphs, as a powerful tool to prove the existence of graphs, have been widely used to gain insight on graph behavior and have been applied more broadly to solve combinatorial problems. One of the problems is to determine how large the probability p needs to be in an Erdős–Rényi graph G(n,p) to ensure the presence of a specific fixed-size subgraph with high probability. In this presentation, we explore the identification of maximum cliques in random graphs and address this problem using some methods from chance-constrained combinatorial programming.

3 - Adaptive Sampling-based Nonconvex and Nonsmooth approaches for Stochastic Programs with Implicitly Decision-dependent Uncertainty

Boyang Shen, Tsinghua University, Beijing, China, People's Republic of, Junyi Liu

We consider a class of stochastic programming problems where the implicitly decision-dependent random variable follows a nonparametric regression model with heteroskedastic error. By the compositional structure, the stochastic optimization problem is weakly convex under suitable conditions. However, the Clarke subdifferential and surrogate functions are not readily obtainable due to the latent decision dependency. To deal with such a computational difficulty, we develop an adaptive sampling-based algorithm that integrates the simulation scheme and statistical estimates to construct sampling-based surrogate functions in a way that the simulation process is guided by the algorithmic procedure. We establish the nonasymptotic convergence analysis in terms of (ϵ, δ)-nearly stationarity in expectation under variable proximal parameters and batch sizes that leads to superior convergence rate and more stable performance in both theory and practice. Furthermore, we show that the proposed adaptive simulation scheme embedded in the sampling-based algorithm leads to better error control of sampling-based surrogate functions and thus further enhance the stability of the sampling-based algorithm. We provide numerical results which illustrate the benefits of variable parameter setting and adaptive simulation schemes in terms of algorithmic stability and efficiency.

4 - Chance Constrained Optimization with Gaussian Mixture Models

Sanjay Mehrotra, Northwestern University, Evanston, IL, United States, Shibshankar Dey, Anirudh Subramanyam, Yi Tianyang, Adrian Maldonado

We develop mixed-integer quadratically constrained program (MIQCP) and second-order cone program (SOCP) formulations for chance-constrained optimization problems with Gaussian mixture models. By leveraging the convex-concave characteristics of the standard normal cumulative distribution function (CDF), we provide both piecewise-linear inner and outer approximations of the original problem. Optimal objectives from these approximations are shown to lie within an epsilon difference of the true value with an adequate number of linear pieces. Additionally, for the conservative SOCP formulation, a bound for relative optimality gap is provided. To expedite computation, we outline a heuristic that selects a significantly lower number of linear pieces than our derived worst-case complexity, without sacrificing desired approximation accuracy. Through extensive numerical experiments analyzing the trade-off between solution time and approximation quality, we find that our formulations offer a significant computational advantage over classical Sample Average Approximation (SAA) methods across various scenarios, including different problem classes and sizes, mixture components, chance constraint reliability levels, and reference distributions.

MC33

Summit - 424

Methods for Large-Scale Nonlinear and Stochastic Optimization IV

Invited Session

OPT: Optimization Under Uncertainty

Chair: Shagun Gupta, UT Austin, Austin

Co-Chair: Jiahao Shi, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Raghu Bollapragada, The University of Texas at Austin, Austin, TX, United States

1 - A Smoothed Augmented Lagrangian Framework for Convex Optimization with Nonsmooth Stochastic Constraints

Peixuan Zhang, Penn State University, University Park, PA, United States, Uday Shanbhag

Motivated by the need to develop simulation optimization methods for more general problem classes, we consider a convex stochastic optimization problem where both the objective and constraints are convex but possibly complicated by uncertainty and nonsmoothness. We present a smoothed sampling-enabled augmented Lagrangian framework that relies on inexact solutions to the AL subproblem, obtainable via a stochastic approximation framework. Under a constant penalty parameter, it is shown that the dual suboptimality diminishes at a rate of $\mathcal{O}(1/K)$ while primal infeasibility and suboptimality both diminish at a rate of $\mathcal{O}(1/\sqrt{K})$. Lastly, we show the flexibility of the framework in accommodating compositional convex constraints and weakly convex objectives.

2 - Modified Line Search Sequential Quadratic Methods for Equality-Constrained Optimization with Unified Global and Local Convergence Guarantees

Raghu Bollapragada, The University of Texas at Austin, Austin, TX, United States

We propose a method that has foundations in the line search sequential quadratic programming paradigm for solving general nonlinear equality constrained optimization problems. The method employs a carefully designed modified line search strategy that utilizes second-order information of both the objective and constraint functions, as required, to mitigate the Maratos effect. Contrary to classical line search sequential quadratic programming methods, our proposed method is endowed with global convergence and local superlinear convergence guarantees. Moreover, we extend the method and analysis to the setting in which the constraint functions are deterministic but the objective function is stochastic or can be represented as a finite-sum. We also design and implement a practical inexact matrix-free variant of the method. Finally, numerical results illustrate the efficiency and efficacy of the method.

3 - Optimization with Highly Adversarial Gradient and Function Value Corruptions: A Unified Framework for High-Probability Iteration Complexity Analysis

Miaolan Xie, Cornell University, Ithaca, NY, United States, Katya Scheinberg

We consider an unconstrained continuous optimization problem in the setting where the gradients can be arbitrarily corrupted in every iteration with a probability greater than $\frac 1 2$. Additionally, the function value estimates may be noisy or adversely corrupted throughout the algorithm's duration. We provide a high probability iteration complexity analysis framework for such an adversarial setting. This analysis offers a unified approach to adaptive stochastic optimization, accommodating noisy or corrupted function value estimates and encompassing both line search and trust region methods.

MC34

Summit - 425

Novel First-order Methods via Performance Estimation Programming

Invited Session

OPT: Linear and Conic Optimization

Chair: Alex Wang, Purdue University, West Lafayette, IN, United States

Co-Chair: Benjamin Grimmer, Johns Hopkins University, 4335 La Plata Ave, Baltimore, MD, 21211, United States

1 -
Toward a Grand Unified Theory of Accelerations in Optimization and Machine Learning

Ernest Ryu, Seoul National University, Department of Mathematical Sciences, Los Angeles, CA, United States

Momentum-based acceleration of first-order optimization methods, first introduced by Nesterov, has been foundational to the theory and practice of large-scale optimization and machine learning. However, finding a fundamental understanding of such acceleration remains a long-standing open problem. In the past few years, several new acceleration mechanisms, distinct from Nesterov’s, have been discovered, and the similarities and dissimilarities among these new acceleration phenomena hint at a promising avenue of attack for the open problem. In this talk, we discuss the envisioned goal of developing a mathematical theory unifying the collection of acceleration mechanisms and the challenges that are to be overcome.

2 - Nonlinear Conjugate Gradient Methods: Worst-case Convergence Rates via Computer-assisted Analyses

Shuvomoy Das Gupta, MIT Operations Research Center, Cambridge, MA, United States, Robert Freund, Andy Sun, Adrien Taylor

We propose a computer-assisted approach to the analysis of the worst-case convergence of nonlinear conjugate gradient methods (NCGMs). Those methods are known for their generally good empirical performances for large-scale optimization, while having relatively incomplete analyses. Using our computer-assisted approach, we establish novel complexity bounds for the Polak-Ribière-Polyak (PRP) and the Fletcher-Reeves (FR) NCGMs for smooth strongly convex minimization. In particular, we construct mathematical proofs that establish the first non-asymptotic convergence bound for FR (which is historically the first developed NCGM), and a much improved non-asymptotic convergence bound for PRP. Additionally, we provide simple adversarial examples on which these methods do not perform better than gradient descent with exact line search, leaving very little room for improvements on the same class of problems.

3 - Novel first-order methods via performance estimation programming

Kevin Shu, Georgia Institute of Technology, Atlanta, GA, United States, Alex L. Wang, Benjamin Grimmer

We will discuss recent work on designing new first order methods via the performance estimation problem. This methodology uses semidefinite programming to analyze the effect of multiple steps of the algorithm in unison, rather than examining each step separately. Our main contribution is showing that it is possible to achieve asymptotic convergence rates of $O(1/T^{1.27})$ for smooth convex functions by just varying the step sizes of gradient descent. If time allows, we may also discuss additional results.

Joint work with Ben Grimmer and Alex Wang.

4 - The Exact Worst-case Convergence Rate of the Alternating Direction Method of Multipliers (ADMM)

Hadi Abbaszadehpeivasti, Tilburg University, Tilburg, Netherlands, Moslem Zamani, Etienne de Klerk

Recently, semidefinite programming performance estimation has been employed as a strong tool for the worst-case performance analysis of first-order methods. In this talk, I derive new non-ergodic convergence rates for the alternating direction method of multipliers (ADMM) by using performance estimation. I also study the linear and R-linear convergence of ADMM in terms of dual objective. I establish that ADMM enjoys a global linear convergence rate if and only if the dual objective satisfies the Polyak–Łojasiewicz (PŁ) inequality in the presence of strong convexity. In addition, I give an explicit formula for the linear convergence rate factor. Moreover, I study the R-linear convergence of ADMM under two scenarios.

MC35

Summit - 427

TSL and TS Best Paper Award

Award Session

Transportation Science and Logistics (TSL)

Chair: Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States

Co-Chair: Karen Smilowitz, Northwestern University, Evanston, IL, United States

1 - Tactical Design of Same-Day Delivery Systems

Alan Erera, Georgia Tech, Atlanta, GA, United States, Alexander Stroh, Alejandro Toriello

coming soon

2 - Charging Station Location and Sizing for Electric Vehicles Under Congestion

Omer Kinay, Amazon, Toronto, ON, Canada, Fatma Gzara, Sibel Alumur

This paper studies the problem of determining the strategic location of charging stations and their capacity levels under stochastic electric vehicle flows and charging times taking into account the route choice response of users. The problem is modeled using bilevel optimization, where the network planner (leader) minimizes the total infrastructure cost of locating and sizing charging stations while ensuring a probabilistic service requirement on the waiting time to charge. Electric vehicle users (followers) minimize route length and may be cooperative or noncooperative. Their choice of route in turn determines the charging demand and waiting times at the charging stations and hence, the need to account for their decisions by the network planner. The bilevel problem reduces to a single-level mixed-integer model using the optimality conditions of the follower's problem when the charging stations operate as M/M/c queues and the followers are cooperative. To solve the bilevel model, a decomposition-based solution methodology is developed that uses a new logic-based Benders algorithm for the location-only problem. Computational experiments are performed on benchmark and real-life highway networks, including a new eastern U.S. network. The impact of route choice response, service requirements, and deviation tolerance on the location and sizing decisions are analyzed. The analysis demonstrates that stringent service requirements increase the capacity levels at open charging stations rather than their number and that solutions allowing higher deviations are less costly. Moreover, the difference between solutions under cooperative and uncooperative route choices is more significant when the deviation tolerance is lower.

3 - How to Deploy Robotic Mobile Fulfillment Systems

René de Koster, Erasmus University, Rotterdam, Netherlands, Lu Zhen, Zheyi Tan, Shuaian Wang

4 - Two-Sided Deep Reinforcement Learning for Dynamic Mobility-on-Demand Management with Mixed Autonomy

Yang Liu, National University of Singapore, Singapore, Singapore, Jiaohong Xie, Nan Chen

coming Soon

5 - Multimodal Vaccine Distribution Network Design with Drones

Shakiba Enayati, University of Missouri - Saint Louis, St Louis, MO, United States, Haitao Li, James Campbell, Deng Pan

Childhood vaccines are essential for public health, yet in remote regions, poor infrastructure and cold chain challenges hinder vaccine availability. This study explores the potential of drones (UAVs) for efficient vaccine delivery by addressing the problem of multimodal vaccine distribution. We develop two mathematical optimization models to strategically determine the locations of distribution centers, drone bases, and relay stations, while respecting cold chain time constraints and drone range limitations. The first model offers a compact formulation that aggregates vaccine travel time but can be overly conservative in adhering to cold chain requirements. The second model uses a layered network framework to track vaccine flow for each origin-destination pair, allowing control over the number of transshipments and drone stops, which improves practical feasibility. Both models are applied to the vaccine distribution network in Vanuatu, using two types of drones as a case study. The results demonstrate significant cost savings with optimal drone usage, highlighting distinct roles for large and small drones

MC36

Summit - 428

Machine Learning/Reinforcement Learning for Vehicle Routing Problems

Invited Session

TSL: Freight Transportation

Chair: Xinwei Chen, Bucknell University, Lewisburg, PA, 74104, United States

1 - The Iterative Chainlet Partitioning Algorithm for the TSP with Drone and Neural Acceleration

Jae Hyeok Lee, KAIST, Daejeon, Korea, Republic of, Minjun Kim, Jinkyoo Park, Changhyun Kwon

This study introduces a novel algorithm, Iterative Chainlet Partitioning (ICP), to solve the Travelling Salesman Problem with Drones (TSP-D). ICP divides TSP-D routes into smaller segments, called chainlets, and applies a precise subroutine, namely TSP-ep-all, to each chainlet. The methodology selects chainlets with the highest incremental improvement for optimization, enabling rapid convergence to high-quality solutions, even for large-scale problems. To further enhance our approach, we integrate a Graph Neural Network (GNN) to predict incremental improvements, significantly reducing the computational burden. Comparative evaluations show that ICP outperforms existing algorithms on most benchmark instances in terms of both solution quality and computational time.

2 - What can we learn from reinforcement learning? Explainable reinforcement learning for on-demand delivery

Xinwei Chen, The University of Tulsa, Tulsa, OK, United States, Xiexin Liu

In recent years, an increasing amount of research has been focused on applying reinforcement learning (RL) to address challenges in on-demand delivery. While various methods have shown improvements over traditional approaches, little attention has been paid to why. This lack of insight into the decision-making process hinders the practical implementation of RL policies, raising concerns about their robustness, etc. In this talk, we introduce a novel framework for learning from RL. We first train the policy for an on-demand delivery problem using an existing deep RL approach. Then, we use supervised learning to learn how different states and actions are mapped. Finally, using the insights we learn from the RL policy, we design policies that are easy to implement and explain to stakeholders. Computational results demonstrate that our proposed policy shows better performance in robustness compared to the conventional RL policy.

3 - Revenue management of customer delivery time offers in on-demand delivery systems

Mathias Klapp, Pontificia Universidad Católica de Chile, Santiago, Chile, Diego Mendieta, Felipe Lagos

We study operational revenue management decisions for on-demand delivery systems by formulating and solving the Personal Shopper Problem with Delivery Time Offers. In this crowdsourced grocery delivery service model, each arriving customer request is offered a delivery time window upon completion of shopping and before payment occurs. Once the offer is made, each customer accepts or rejects it based on their personal preferences. If the customer accepts the offer, the request enters the system and must be served. If the customer rejects the offer, the request is lost, and the platform loses revenue. We design decision policies via an ad-hoc value function approximation that minimizes expected lost revenue and customer dissatisfaction.

4 - Neural Genetic Operators for the Traveling Salesman Problem

Hyeonah Kim, KAIST, Daejeon, Korea, Republic of, Jae Hyeok Lee, Jinkyoo Park, Changhyun Kwon

Genetic algorithms are widely recognized as effective meta-heuristics for tackling diverse combinatorial optimization, including the traveling salesman problem. However, the performance of genetic algorithms heavily relies on the careful design for genetic operations, i.e., crossover, mutation, and selection, leading to the requirements of deep comprehension and problem-specific knowledge. Consequently, genetic algorithms face challenges in directly applying to newly emerging problems that have been insufficiently studied. To mitigate the lack of task expandability, we propose an algorithm design framework for genetic algorithms utilizing the recent progress in deep learning. This allows a general design process of genetic operators for various routing problems.

5 - Serving E-Commerce and E-Grocery Deliveries Together with Learning-Based Vehicle Dispatching

Gürhan Kok, Koc University, Istanbul, Turkey, Oguz Sohret, Barış Yıldız

We examine a collaboration process between a grocery and an e-commerce company, who agree to deliver online orders using a single vehicle. Grocery orders including same-day-delivery are delivered within time windows, whereas e-commerce orders can be delivered anytime within a day.

We present a machine-learning based approach to support the routing optimization model and compare it with reference models.

MC37

Summit - 429

Emerging Technologies (AVs, EVs, robots, lockers, etc.)

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Tho Le, Purdue University, West Lafayette, IN, United States

1 - A Two-Stage Stochastic Optimization Model for Sidewalk Robot Food Delivery Systems

Tho Le, Purdue University, West Lafayette, IN, United States

The rapidly expanding online food delivery (OFD) market poses significant challenges to last-mile delivery operations. Sidewalk delivery robots (SDRs) offer a promising alternative to relying on on-demand workers, as these box-sized robots can quickly deliver food or groceries within a short range via sidewalks. This paper proposes a two-stage stochastic optimization model for a multi-depot SDR system. The first stage determines the optimal fleet size for each depot and the allocation of battery-swapping facilities at some of the depots. The second stage solves the routing problem with a late arrival penalty. A novel solution algorithm is developed, utilizing continuous approximation (CA) in the first stage to quickly obtain a reduced solution space, and a heuristic algorithm in the second stage to provide high-quality solutions efficiently. The CA model considers factors such as time constraints, battery swapping, and order compliance. The second-stage results guide the adjustment of the CA model used in the first stage. Finally, the sample average approximation (SAA) method is applied to choose the solution corresponding to the lowest average second-stage objective value. A numerical example demonstrates the effectiveness of the proposed approach in reducing computational time while maintaining accuracy. Future work includes evaluating the proposed approach against benchmarks and solving large-scale instances.

2 - Deploying Autonomous Mobile Parcel Locker in Last-mile Parcel Delivery

Yubin Liu, Singapore-MIT Alliance for Research and Technology Center, Singapore, Singapore, Dingtong Yang, Hamsa Balakrishnan, Hai Wang

The vigorous growth of e-commerce transactions has intensified the need for rapid and adaptable last-mile delivery services. Conventional delivery methods involving human-operated vehicles may not adequately meet increasing demands due to escalating labor costs and a scarcity of trained personnel. In response, Mobile Parcel Lockers (MPLs), supported by advancements in autonomous driving technology, have been increasingly implemented by logistics providers as a viable solution to these challenges. Compared to static parcel lockers, MPLs capitalize on their mobility advantage to reduce the walking distance required for customers to retrieve/deliver their packages, thereby enhancing customer convenience. This research investigates a scenario where MPLs dynamically relocate their positions according to customers' spatial distribution to optimize parcel pick-up and delivery processes, considering factors such as customer walking distances and preferred service time windows. The study introduces a mixed integer programming model to optimize the operations of MPLs, aiming to reduce operational costs. Specifically, the research seeks to determine i) the optimal locations for MPL stopovers to maximize the customers being served at each stopover, ii) the optimal duration of service at each MPL stopover, and iii) the most efficient routing for MPL fleets among various stopovers. By examining the impact of customer walking distances on the stopover positioning, the service duration of each stopover, and operational costs, this study assists operators in more effectively allocating MPL services. This strategic deployment allows operators to meet customer needs more cost-effectively, thereby enhancing service reach and efficiency.

3 - Vehicle Routing Problems in the Age of Semi-Autonomous Driving

Hins Hu, Cornell University, Ithaca, NY, United States, Samitha Samaranayake

We are in the midst of a semi-autonomous era in urban transportation in which varying forms of vehicle autonomy are gradually being introduced. This phase of partial autonomy is anticipated by some to span a few decades due to various challenges, including budgetary constraints to upgrade the infrastructure and technological obstacles in the deployment of fully autonomous vehicles (AV) at scale. In this study, we introduce the vehicle routing problem in a semi-autonomous environment (VRP-SA) where the road network is not fully AV-enabled in the sense that a portion of it either not suitable for AVs or requires additional resources in real-time (e.g., remote control) for them to pass through. Moreover, such resources are scarce and usually subject to a budget constraint. An exact mixed-integer linear program (MILP) is formulated to minimize the total routing cost of service in this environment. We propose a two-phase algorithm based on a family of feasibility recovering sub-problems to solve the VRP-SA efficiently. Our algorithm is implemented and tested on a new set of instances that are tailored for the VRP-SA by adding stratified grid road networks to the benchmark instances. The result demonstrates a reduction of up to 37.5% in vehicle routing costs if the fleet actively exploits the AV-enabled roads in the environment. Additional analysis reveals that cost reduction is higher with more budget and longer operational hours.

4 - Collaborative Electric Vehicle Routing for Energy Consumption Minimization

Fangting Zhou, Chalmers University of Technology, Gothenburg, Sweden, Ivan Sanchez-Diaz, Balazs Kulcsar

Electric vehicles for urban delivery have emerged as a promising solution to mitigate the environmental impact. This paper proposes a collaborative electric vehicle routing problem (CoEVRP) for urban logistics, in which the exchange of goods at meet points is strategically planned en route. The objective of the proposed CoEVRP is to minimize energy consumption. Time and space synchronization constraints are considered for better collaboration between vehicles. An energy estimation model is introduced to estimate energy consumption in the routes by considering the variation of ambient temperature. The proposed problem is modeled as a mixed-integer linear programming problem. This paper develops a matheuristic method that combines adaptive large neighborhood search with linear programming for real-world implementations. The study is demonstrated using the real case of two grocery stores in the city of Gothenburg and large-scale experiments. The findings show that on top of reducing travel distance and costs, horizontal collaboration between transportation companies is environmentally sustainable, with around 20% less energy consumption. This paper also suggests that the collaboration between vehicles is more significant during the winter season when the battery performance is more subjected to low temperatures.

MC38

Summit - 430

Competition in the Aviation Industry

Invited Session

Aviation Applications

Chair: Gianmarco Andreana, University of Bergamo, Bergamo, Italy

1 - Demand Estimates and Valuing Slots

Gianmarco Andreana, University of Bergamo, Bergamo, Italy

This paper estimates air transport demand for business and leisure segments in the European, North American, and transatlantic markets, with two primary objectives. First, we estimate a set of general logit demand parameters for fares, frequencies, distances, and directness, which can be used in applied (game) theoretical models. Second, we examine the differences in demand parameters across the various segments and markets and investigate the implications of a slot reduction at two major hubs, one in Europe (London) and one in North America (Atlanta). We explore how these effects propagate through the network. Our game-theoretic model of airline competition with slot constraints at these major hubs reveals non-trivial outcomes, including an increase in airline market power, a reduction in consumer surplus, and a decrease in environmental damage. These findings highlight the sophisticated impacts of slot regulation on the aviation industry and demonstrate the importance of considering the effects of slot reductions at multiple major hubs across different regions.

2 - Two-stage Drone Delivery Network Design with Uncertain Demand

Wenjia ZENG, Hong Kong University of Science & Technology, Hong Kong, China, People's Republic of, Ruiwei Jiang, Hai Yang, Hai Wang

Drones are expected to play an important role in the logistics and transportation sectors, with numerous nations and corporations already initiating pilot programs for the rapid delivery of goods, including fast food and groceries, using drones. There is a burgeoning interest in the potential for broader application of this drone delivery service. In light of this, this paper studies a drone delivery network design problem with uncertain demand, particularly focusing on the urban low-altitude context. Our goal is to achieve a cost-effective solution that minimizes both the investment costs on network infrastructure and the subsequent operational costs. A critical aspect of our paper is the consideration of both the soft and hard time windows for customers and the capacity of drone parking and recharging facilities. The lack of extensive historical data on demand patterns, coupled with inherent variability in consumer behavior and market dynamics, also poses challenges for our problem. To tackle this challenge, we propose a two-stage distributionally robust optimization approach, which utilizes a distribution separation procedure and split cuts within Benders decomposition algorithm. The validity and robustness of our proposed method are demonstrated through extensive experiments.

3 - Impact of Traffic-Following on Order of Autonomous Airspace Operations

Anahita Jain, The University of Texas at Austin, Austin, TX, United States

In air traffic management, fostering safe, expeditious, and orderly traffic are key objectives. Air traffic controllers ensure safe separation distances between vehicles and often expedite traffic to enhance system throughput. In maintaining the orderliness of traffic, controllers ensure aircraft compliance with established route structures, and they apply equitable first-come, first-serve service. They dynamically organize traffic into patterns to effectively navigate complexity and workload challenges, especially in airspaces with high traffic densities. Order is often traded for expediency and safety, for example, deviating from route structures to gain time or resolve traffic conflicts.

In this work, we investigate the dynamic emergence of traffic order in a distributed multi-agent system, aiming to minimize inefficiencies stemming from unnecessary structural impositions. We introduce a methodology for developing a dynamically updating traffic pattern map of the airspace by leveraging information about the consistency and frequency of flow directions used by current and preceding traffic. Informed by this map, an agent can discern the degree to which it is advantageous to follow traffic by trading off utilities such as time and order. We show that for the traffic levels studied, for low degrees of traffic-following behavior, there is minimal penalty in terms of aircraft travel times while improving the overall orderliness of the airspace. On the other hand, heightened traffic-following behavior may result in increased aircraft travel times, while marginally reducing the overall entropy
of the airspace. Ultimately, this work can be used to optimally and dynamically adjust an agent's traffic-following behavior based on these trade-offs.

4 - Exploring the Role of Drones and UAVs in Logistics and Supply Chain Management: A Novel Text-Based Literature Review

Bahareh Kargar, New Jersey Institute of Technolog, Newark, NJ, United States

Drones and Unmanned Aerial Vehicles (UAVs) have become pivotal in Logistics and Supply Chain (LSC) management. Our study uncovers a significant disparity highlighted through an examination of recent review papers, indicating a deficiency in UAV applications across entire supply chains. We utilize a novel text-mining method, incorporating qualitative assessments and temporal trend analysis on 5,364 papers. By employing the Latent Dirichlet Allocation (LDA) model, we identify ten distinct research topics, suggesting four promising avenues for future exploration, including sustainability, scalability, pandemic response, and integration with multimodal transportation, providing insights for advancing drone-based logistics systems.

MC39

Summit - 431

RAS Round Table on Railway Service Reliability - Part 1

Invited Session

Railway Applications

Chair: Michael Gorman, University of Dayton, 300 college park, Dayton, OH, 45419, United States

Co-Chair: Erick Wikum, Wikalytics LLC, 4906 Whispering Creek Ct, Maineville, OH, 45039, United States

1 - Setting the table: Rail Reliability

Michael Gorman, University of Dayton, Dayton, OH, United States

I kickoff the session with an overview of rail reliability and discuss its causes and overview the cures.

2 - Service Design and Rail Reliability

Carl Van Dyke, TransNetOpt, West Windsor, NJ, United States

Reliability and service quality have long been issues in the rail industry. While originally associated primarily with the carload element of the railroad's lines of business, they also appear within the context of both unit train and intermodal operations. The presenter will review the history of the reliability issue, some of the underlying causes, and the possible ways that these issues could be addressed.

3 - Reliability from a Yard Perspective

Edwin Kraft, TEMS, Inc., Frederick, MD, United States

For many years, yards have been a stubborn source of rail shipment delays and missed connections, leading to service failures and poor equipment utilization. However, it can be difficult in yards to physically access arriving cars that need to make specific connections without also processing many other cars that might not be needed or wanted. Variability in train arrival times and traffic volume fluctuations challenge yard managers, while congestion, switch crew capabilities and restrictive track layouts limit management flexibility and options. We will discuss possible operational and infrastructure approaches for making yards more productive, predictable, and controllable given the practical constraints of day-to-day operations.

4 - Business Imperative to Address Service & Reliability

Sanford Sexhus, Oliver Wyman, Dallas, TX, United States

If railroads simply maintain the status quo with respect to service and reliability they are likely to continue to lose market share in the N.A. transport market. If railroads were to meaningfully move the needle with respect to service, then their ability to maintain or increase market share improves, with large, long term, financial rewards. The nature and scale of this opportunity will be presented.

5 - Incorporating reliability into railway operations planning: select real-life applications.

Stefano Rieppi, Norfolk Southern Railway, Roswell, GA, United States

In this presentation, select case studies regarding the incorporation of the reliability concept into railway operations planning will be presented. The area of focus will be rolling stock distribution planning, where variation in reliability can impact timing of availability of rolling stock, as well as progress of said rolling stock towards planned points of loading. The presentation will cover potential failure modes we seek to manage, as well as systems-based processes in place to manage reliability challenges in the context of optimized distribution operations.

6 - Service Reliability and Amtrak

Jeremiah Dirnberger, Amtrak, Jacksonville, FL, United States

Amtrak has three Service Lines: Northeast Corridor, State Supported and Long Distance. I will discuss the various interactions with the freight railroads for each of the Service Lines. These interactions determine the different performance targets for each Service Line and the overall perspective that Amtrak has on Service Reliability.

MC40

Summit - 432

Sustainability and Donations in Behavioral Operations

Invited Session

Behavioral Operations Management

Chair: Leon Valdes, University of Pittsburgh, Pittsburgh, PA, United States

Co-Chair: Gabriel Alfonso Pensamiento Calderon, University of Pittsburgh, Pittsburgh, PA, 15213, United States

1 - Building Alliances for Corporate Social Responsibility

Han Zhang, Michigan State University, East Lansing, MI, United States, Ruth Beer, Kyle Cattani

Problem definition: When multiple companies share the same social responsibility problem, forming an industry alliance can help the companies pool resources, spread the cost, and have greater influence. Companies, however, have the incentive to free ride on the effort of others. What mechanism can avert free riding and let companies form a successful alliance?

Methodology/results: Motivated by an industry case, we model alliance formation as a two-stage public goods game among companies with different brand values and degrees of prosociality. One company can invite another to form an initial alliance before other companies decide to join. We find that the invitation mechanism is the key to success. We identify two mechanisms that affect the type of alliance formed: status seeking and prosocial behavior. We test the theory in an incentivized laboratory experiment. The experiment confirms the essential role of invitation. The experiment also shows that the initiator predominantly invites a high-brand-value company—evidence of status seeking. Once an initial alliance is formed, the prosocial tendency drives further contribution of other high-brand-value companies but not that of low-brand-value companies.

Managerial implications: Establishing an early alliance is the key to addressing a shared social responsibility problem. Companies should leverage status seeking and prosocial tendencies to engage other companies. While there may be clear advantages to inviting companies with highly valued brands, our results show that when the prosocial tendency is strong and status seeking weak, an initiating company may benefit by proactively inviting lesser-known brands early on.

2 - Generic Drug Transparency - Testing a Regulatory Policy Proposal

Sebastian Villa, University of New Mexico, Albuquerque, NM, United States, Gloria Urrea, George Ball, John Gray

The generic drug industry is opaque. Consumers have little ability to determine the quality or manufacturing location of their drugs. We ran two experimental studies to test public's and pharmacists' reaction to increasing generic drug transparency. Our results lead to actionable policy recommendations that can help strengthen the drug supply chain.

3 - On the Impact of Food Packaging On the U.S. Households' Food Waste

Chao Wu, Wayne state university, Rochester Hills, MI, United States, Monireh Mahmoudi, Stanley Lim

In this study, we explore the impact of packaging on U.S. households' food waste. First, we conduct an 8-week diary study with Qualtrics Panelists to examine the impact of various packaging features on household food waste. Then, we confirm the findings by running a follow-up experiment with student population and further explore the potential mechanisms. Implications for operations and policy are discussed.

4 - Reference-dependent effects in volunteering: Expected vs. experienced effort

Leon Valdes, University of Pittsburgh, Pittsburgh, PA, United States, Mahyar Eftekhar, Gabriel Pensamiento

Humanitarian and nonprofit organizations often rely on volunteer labor to accomplish different tasks. We design an incentivized laboratory experiment to study how someone's volunteering experience being above or below effort-expectations affects their decision to volunteer again. In doing so, we study the role of reference-dependent preferences in a not-for-profit setting, which is not well understood in the literature. We also aim to shed light on which responses may be unique to (or more pronounced in) a volunteering setting by comparing our results against a similar task in a for-profit context.

MC41

Summit - 433

Online Learning in Operations

Invited Session

Applied Probability Society

Chair: Sean Sinclair, Northwestern University, Evanston, IL, United States

1 - Multi-Task Contextual Dynamic Pricing

Jingwei Ji, University of Southern California, Los Angeles, CA, United States, Renyuan Xu, Adel Javanmard

We study the dynamic pricing problem faced by a firm which sells a large number of products.
A product is fully characterized by feature, a d dimensional vector.
The customer will buy the product if our price is better than the competitors', and otherwise not.
We assume a structured pricing model of the competitor, whose parameters are unknown a priori.
The parameters for pricing different products may or may not be similar to each other.
We characterize the overall similarity by $\deltamax$.
The firm's objective is to minimize the expected regret, namely, the expected revenue loss against a clairvoyant policy which has the knowledge of the parameters used by the competitor for different products.
We propose a multi-task learning dynamic pricing strategy which leverages the potential similar structure shared by different products.
More specifically, we show that the regret of our policy is better than both a policy which treats each product individually, and a policy which treats all products as the same.
Moreover, the regret is bounded by $\tilde{O} \paren{ \deltamax \sqrt{TM} d } $, where M is the number of products.

2 - Experimental Design in Supply Chains

Yunhong Wang, National University of Singapore, Singapore, Singapore, Jinglong Zhao, Hanzhang Qin

In this paper, we develop a framework for analyzing biases when implementing experiments in a supply chain system. Operations, such as replacing a fast-moving product with a slow-moving product, could have a long-term effect on the demand. The biases occur due to the spillover effect of the supply chain system, as customers who have not fulfilled their needs at the current FC may turn to other FC so there is interference among customers. We build up a stochastic model for this system and use the mean-field limit method to capture system dynamics which depend on local stock-out events and customer behavior. We also specify the source of interference and discuss the trade-off between different types of interference. Finally, we propose some experimental designs and associated estimators and show that our methods can help to reduce biases compared to the classical demand side "customer randomization(CR)" experimental design.

3 - Reinforcement Learning in Mdps with Exogenous Inputs

Sean Sinclair, Northwestern University, Evanston, IL, United States

Many operations management problems require sequential decision-making under uncertainty, where the only uncertainties affecting the outcomes are exogenous variables outside the control of the decision-maker. We introduce a modeling framework known as Markov Decision Processes with Exogenous Inputs (Exo-MDPs), which explicitly incorporates these uncontrollable inputs into the system dynamics. We outline two algorithmic strategies tailored for these models. The first strategy, Hindsight Learning (HL), achieves data efficiency by leveraging a key insight: having samples of the exogenous variables allows past decisions to be revisited in hindsight to infer counterfactual consequences, thereby accelerating policy improvements. The second strategy addresses settings where the exogenous inputs are censored, designing algorithms that exploit low-rank structure to infer the underlying exogenous randomness for data-efficient learning. We compare our algorithms against classic baselines in applications such as cloud computing, airline revenue management, and inventory problems with lost sales and positive lead times.

4 - A Minibatch-Sgd-Based Learning Meta-Policy for Inventory Systems with Myopic Optimal Policy

Shilin Yuan, Tsinghua University, Beijing, China, People's Republic of, Jiameng Lyu, Jinxing Xie, Yuan Zhou

Stochastic gradient descent (SGD) has proven effective in solving many inventory control problems with demand learning. However, it often faces the pitfall of an infeasible target inventory level that is lower than the current inventory level. Several recent works (e.g., (Huh and Rusmevichientong 2009, Shi et al. 2016)) are successful to resolve this issue in various inventory systems. However, their techniques are rather sophisticated and difficult to be applied to more complicated scenarios such as multi-product and multi-constraint inventory systems.

In this paper, we address the infeasible-target-inventory-level issue from a new technical perspective -- we propose a novel minibatch-SGD-based meta-policy. Our meta-policy is flexible enough to be applied to a general inventory systems framework covering a wide range of inventory management problems with myopic clairvoyant optimal policy. By devising the optimal mini-batch scheme, our meta-policy achieves O(T^(1/2)) regret for the general convex case and O(\log T) regret for the strongly convex case. To demonstrate the power and flexibility of our meta-policy, we apply it to three important inventory control problems: multi-product and multi-constraint systems, multi-echelon serial systems, and one-warehouse and multi-store systems by carefully designing application-specific subroutines. We also conduct extensive numerical experiments to demonstrate that our meta-policy enjoys competitive regret performance, high computational efficiency, and low variances among a wide range of applications.

MC42

Summit - 434

Advances in Adaptive Experimentation

Invited Session

Applied Probability Society

Chair: Lin Fan, Northwestern University, Kellogg School of Management, Evanston, IL, United States

1 - Asymptotically Optimal Adaptive A/B Tests for Average Treatment Effect

Vikas Deep, Northwestern University, Evanston, IL, United States, Achal Bassamboo, Sandeep Juneja

Typically in A/B testing, an experiment designer sequentially assigns treatment A or B to arriving individuals to identify the better treatment (known as the best treatment identification (BTI) problem), that is, the one with better mean performance. We focus on a related, equally important, but more informative problem of estimating the difference between the two means, defined as the average treatment effect (ATE). For computational efficiency, we restrict accuracy to a confidence interval (CI) of width at most ε, where the probability of CI not containing ATE is restricted to at most δ. The objective is to estimate a CI of ATE with the minimum expected sample size. We first establish a lower bound on the expected sample size of the A/B test needed for any adaptive policy, which constructs a CI of ATE with desired properties as the

solution to a max-min problem for small δ. The min-max problem provides the asymptotic fraction of the assignment of treatments A and B for any asymptotically optimal policy.
Using the insights provided by the max-min problem, we construct an adaptive policy that is asymptotically optimal, i.e., matches the lower bound on the expected sample size for small ε and δ. Finally, we present a comparative analysis between our ATE problem and the BTI problem revealing marked differences in the asymptotically optimal fraction of assignment of treatments in ATE and BTI problems.

2 - Optimizing Adaptive Experiments: A Unified Approach to Regret Minimization and Best-Arm Identification

Chao Qin, Stanford University, Stanford, CA, United States, Daniel Russo

Practitioners conducting adaptive experiments often encounter two competing priorities: maximizing total welfare (or `reward') through effective treatment assignment and swiftly concluding experiments to implement population-wide treatments. Current literature addresses these priorities separately, with regret minimization studies focusing on the former and best-arm identification research on the latter. This paper bridges this divide by proposing a unified model that simultaneously accounts for within-experiment performance and post-experiment outcomes. We provide a sharp theory of optimal performance in large populations that not only unifies canonical results in the literature but also uncovers novel insights. Our theory reveals that familiar algorithms, such as the recently proposed top-two Thompson sampling algorithm, can optimize a broad class of objectives if a single scalar parameter is appropriately adjusted. In addition, we demonstrate that substantial reductions in experiment duration can often be achieved with minimal impact on both within-experiment and post-experiment regret.

3 - Brownian Bandits

Dawei Li, University of Chicago, Chicago, IL, United States, Baris Ata, Michael Harrison

We consider a Gaussian bandit formulation in continuous time and solve the corresponding Bellman equation using deep learning. Leveraging this solution, we propose policies for the discrete-time analogs of our formulation and provide comparisons with benchmark policies.

4 - Robustness Benefits of Structured Bandits

Lin Fan, Northwestern University, Evanston, IL, United States, Peter Glynn

We study the tail of the regret distribution of algorithms optimized for asymptotic expected regret in structured bandit settings. In classical bandit settings, the rewards collected for an arm do not provide information about the quality of the other arms. In the absence of such information-sharing, optimized algorithms quite often make catastrophic statistical errors in estimating the quality of the arms, thereby resulting in very heavy regret distribution tails and fragility to model mis-specification, as exposed in Fan and Glynn (2021). By contrast, in structured bandit settings with information-sharing among arms, we show that optimized algorithms can be significantly more robust.

MC43

Summit - 435

Robust MDPs and Reinforcement Learning

Invited Session

Applied Probability Society

Chair: Yashaswini Murthy, University of Illinois Urbana-Champaign, Urbana, IL, 61820, United States

1 - On the Convergence of Modified Policy Iteration in Exponential Cost Risk Sensitive MDPs

Yashaswini Murthy, University of Illinois Urbana-Champaign, Urbana, IL, United States

Modified Policy Iteration (MPI) is a dynamic programming algorithm that combines elements of policy iteration and value iteration. The convergence of MPI has been well studied in the context of discounted and average-cost MDPs. In this work, we consider the exponential cost risk-sensitive MDP formulation, which is known to provide robustness to model parameters. Although policy iteration and value iteration have been well studied in the context of risk sensitive MDPs, MPI is unexplored. We provide the first proof that MPI also converges for the risk-sensitive problem in the case of finite state and action spaces. Since the exponential cost formulation deals with the multiplicative Bellman equation, our main contribution is a convergence proof which is quite different from the proof of existing results for discounted and risk-neutral average-cost problems as well as risk sensitive value and policy iteration approaches. We conclude our analysis with simulation results, assessing MPI's performance relative to alternative dynamic programming methods like value iteration and policy iteration across diverse problem parameters. Our findings highlight risk-sensitive MPI's enhanced computational efficiency compared to both value and policy iteration techniques.

2 - Achieving the minimax optimal sample complexity of offline reinforcement learning: a DRO-based approach

Shaofeng Zou, Arizona State University, Tempe, AZ, United States, Yue Wang, Jinjun Xiong

Offline reinforcement learning aims to learn from pre-collected datasets without active exploration. This problem faces significant challenges, including limited data availability and distributional shifts. Existing approaches adopt a pessimistic stance towards uncertainty by penalizing rewards of under-explored state-action pairs to estimate value functions conservatively. In this paper, we show that the distributionally robust optimization (DRO) based approach can also address these challenges and is \textcolor{blue}{asymptotically minimax optimal}. Specifically, we directly model the uncertainty in the transition kernel and construct an uncertainty set of statistically plausible transition kernels. We then show that the policy that optimizes the worst-case performance over this uncertainty set has a near-optimal performance in the underlying problem. We first design a metric-based Hoeffding-style uncertainty set such that with high probability the true transition kernel is in this set. We prove that to achieve a sub-optimality gap of ϵ, the sample complexity is $\mathcal{O}(S^2C^{\pi^*}\epsilon^{-2}(1-\gamma)^{-4})$, where γ is the discount factor, S is the number of states, and C^{π^*} is the single-policy clipped concentrability coefficient which quantifies the distribution shift. To achieve the optimal sample complexity, we further propose a less conservative Bernstein-style uncertainty set, which, however, does not necessarily include the true transition kernel. We show that an improved sample complexity of $\mathcal{O}(SC^{\pi^*}\epsilon^{-2}(1-\gamma)^{-3})$ can be obtained, which asymptotically matches with the minimax lower bound for offline reinforcement learning, and thus is asymptotically minimax optimal.

3 - Policy Evaluation for Variance in Average Reward Reinforcement Learning

Shubhada Agrawal, Georgia Institute of Technology, Atlanta, GA, United States

We consider an average reward reinforcement learning (RL) problem and work with asymptotic variance as a risk measure to model safety-critical applications. We design a temporal-difference (TD) type algorithm tailored for policy evaluation in this context. Our algorithm is based on linear stochastic approximation of an equivalent formulation of the asymptotic variance in terms of the solution of the Poisson equation. We consider both the tabular and linear function approximation settings, and establish $\tilde {O}(1/k)$ finite time convergence rate, where k is the number of steps of the algorithm. Our work paves the way for developing actor-critic style algorithms for variance-constrained RL. To the best of our knowledge, our result provides the first sequential estimator for asymptotic variance of a Markov chain with provable finite sample guarantees, which is of independent interest.

4 - First-order Policy Optimization for Robust Policy Evaluation

Guanghui Lan, Georgia Institute of Technology, Atlanta, GA, United States, Yan Li

We adopt a policy optimization viewpoint towards policy evaluation for robust Markov decision process with s-rectangular ambiguity sets. The developed method, named first-order policy evaluation (FRPE), provides the first unified framework for robust policy evaluation in both deterministic (offline) and stochastic (online) settings, with either tabular representation or generic function approximation. In particular, we establish linear convergence in the deterministic setting, and O(1/\epsilon^2) sample complexity in the stochastic setting. FRPE also extends naturally to evaluating the robust state-action value function with (s,a)-rectangular ambiguity sets. We discuss the application of the developed results for stochastic policy optimization of large-scale robust MDPs.

5 - Prelimit Coupling and Steady-State Convergence of Constant-stepsize Nonsmooth Contractive SA

Yixuan Zhang, University of Wisconsin-Madison, Madison, WI, United States, Lucy Huo, Yudong Chen, Qiaomin Xie

Motivated by Q-learning, we study nonsmooth contractive stochastic approximation (SA) with constant stepsize. We focus on two important classes of dynamics: 1) nonsmooth contractive SA with additive noise, and 2) synchronous and asynchronous Q-learning, which features both additive and multiplicative noise. For both dynamics, we establish weak convergence of the iterates to a stationary limit distribution in Wasserstein distance. Furthermore, we propose a prelimit coupling technique for establishing steady-state convergence and characterize the limit of the stationary distribution as the stepsize goes to zero. Using this result, we derive that the asymptotic bias of nonsmooth SA is proportional to the square root of the stepsize, which stands in sharp contrast to smooth SA. This bias characterization allows for the use of Richardson-Romberg extrapolation for bias reduction in nonsmooth SA.

MC44

Summit - 436

Large-Scale Simulation Optimization

Invited Session

Simulation Society

Chair: Jinbo Zhao, Texas A&M Univeristy, College Station, TX, United States

Co-Chair: David Eckman, Texas A&M University, College Station, TX, United States

1 - Using Functional Properties to Screen Solutions for Multi-Objective Simulation Optimization

Jinbo Zhao, Texas A&M Univeristy, College Station, TX, United States, David Eckman

We propose a class of screening procedures designed for multi-objective simulation-optimization problems. The procedures screen out solutions that are deemed Pareto-dominated by leveraging known or assumed properties of the functions describing solution performance, such as convexity and Lipschitz continuity, while simulating only a small subset of all solutions. Whether to screen out a given solution is determined by solving a non-convex mixed-integer program, the size of which scales with the number of simulated solutions and the number of objectives. The screening decisions are accompanied by several statistical guarantees, including uniform confidence and consistency. We discuss how modeling choices for the underlying optimization problem impact both the statistical effectiveness and the computational efficiency of the procedures and demonstrate the resulting performance via numerical experiments. Extensions to problems with stochastic constraints will also be discussed.

2 - Dice and Slice Simulation Optimization for High-Dimensional Problems

Harun Avci, Northwestern University, Evanston, IL, United States

Stochastic optimization problems associated with real-life applications sometimes can be simplified enough to get managerial insights, or structured enough to solve via mathematical optimization, but frequently simulation optimization (SO) is needed. We describe such a problem arising from our collaboration with General Motors on “future mobility.” Unfortunately, large-scale problems with high-dimensional discrete solution spaces are computationally impossible to solve with current SO technology. To make solving such problems computationally feasible, we propose the Dice and Slice Simulation Optimization (DASSO) methodology and algorithm. This method decomposes a Bayesian prior distribution on the objective function into an additive form, reducing the problem dimensionality to facilitate efficient posterior updates, and identifies the most-promising solutions efficiently. Our results reveal that it can find good solutions rapidly on large-scale problems with more than a trillion feasible solutions, far beyond the reach of any other discrete SO algorithm.

3 - Sample-Efficient Clustering and Conquer Procedures for Parallel Large-Scale Ranking and Selection

ZISHI ZHANG, Peking University, Beijing, China, People's Republic of, Yijie Peng

We propose novel "clustering and conquer" procedures for the parallel large-scale ranking and selection (R&S) problem, which leverage correlation information for clustering to break the bottleneck of sample efficiency. In parallel computing environments, correlation-based clustering can achieve an O(p) sample complexity reduction rate, which is the optimal reduction rate theoretically attainable. Our proposed framework can achieve improvements without the necessity of highly accurate correlation estimation and precise clustering. Additionally, we propose a parallel few-shot clustering algorithm tailored for large-scale problems. In large-scale AI applications such as neural architecture search, our method demonstrates excellent numerical performance.

4 - Evaluating Solvers for Linearly Constrained Simulation Optimization

Litong Liu, Georgia Institute of Technology, Atlanta, GA, United States

Linearly constrained simulation optimization problems are those that include deterministic linear constraints in addition to an objective function that can only be evaluated through simulation. We provide several solvers for linearly constrained simulation optimization that all rely on gradient estimates of the objective function. We compare these solvers on random instances of 4 test problems from SimOpt.

MC45

Summit - 437

Optimizing Hospital and Public Health Operations

Invited Session

Health Applications Society

Chair: Vikram Tiwari, Vanderbilt University Medical Center, Nashville, TN, United States

1 - Data-Driven Approach to the Surgical Block Auto-Release Decision

Sandeep Rath, University of North Carolina at Chapel Hill - Kenan Flagler, Chapel Hill, NC, United States, Moe Lim

Surgical block allocation involves reserving block time in operating rooms (OR) for surgeon scheduling. This gives the surgeons independence to manage their own schedules predictably and ensures fair distribution of OR capacity across departments. However, unused block time can result in idle time in the OR, leading to lower OR utilization. In order to improve utilization and total case volume, unused blocks are auto-released for open access several days before surgery. Deciding how far in advance to auto-release block time is a trade-off between predictable surgeon scheduling and OR utilization. Although block allocation optimization is a well-studied topic, the decision on how far in advance to automatically release unfilled block time has received less attention in the literature. To address this gap, we developed a data-driven approach for surgical block auto-release using historical data of surgical case bookings timestamps from a large academic medical center. Our model improves OR utilization without compromising predictable surgeon scheduling.

2 - Machine Learning from Social Determinants of Health in Predicting Diabetes Outcomes

Seung-Yup Lee, University of Alabama at Birmingham, Birmingham, AL, United States, Musawir Haseeb, Mohammad Saleem, Leslie Hayes, Alison Garretson, Heather Bradley, Erin DeLaney, Andrew Land, Andrea Cherrington, Carrie Howell, Allyson Hall, Jane Banaszak-Holl

Despite the increasing recognition of social determinants of health (SDoH) as key factors in diabetes patient health trajectories, their integration into predictive models for tailored interventions remains limited.

In this study, we aim to employ large-scale SDoH data to predict outcomes for diabetes patients. Our retrospective prediction modeling study utilizes Latent Dirichlet Allocation for natural language processing (NLP) topic modeling and the Least Absolute Shrinkage and Selection Operator (LASSO) for feature selection. These methods analyze a comprehensive set of over 100 distinct SDoH items, derived from both categorical and free-text social histories, to develop predictive models for 30-day emergency department (ED) revisits among diabetic patients.

To demonstrate the value of integrating SDoH into prediction modeling, we adopted a progressive model development approach by developing three classification models with increasing sizes of predictors and complexity. Model 1 includes demographic information and geographical SDoH such as the area deprivation index and social vulnerability indices. Model 2 adds 43 categorical SDoH items to Model 1. Model 3 incorporates an additional 59 individual-level SDoH items from free-text data.

We identified 11,775 ED patient-visits to the University of Alabama at Birmingham Healthcare System due to diabetic concerns in 2018. We observed significant improvements across the models: sensitivity of classification increased from 0.098 in Model 1 to 0.332 in Model 3; precision improved from 0.410 to 0.507; and the area under the curve from 0.609 to 0.705. Our results demonstrate the significant predictive capability of SDoH data in forecasting ED revisits among diabetes patients.

3 - Locations of Drone Ambulance Stations to Improve Response TIME of Cardiac Arrest Patients: Ambulance Dynamics in the Euclidean Plane

EUNBI KIM, Korea University, Seoul, Korea, Republic of, Joonyup Eun, Taesu Cheong

This study aims to identify the optimal locations of manned drone ambulance stations to reduce response time for cardiac arrest patients. When responding to a cardiac arrest incident, it is advisable to dispatch either a ground ambulance or a manned drone ambulance, depending on the time of each option taken to reach the patient. The ambulance that can reach the patient in earlier is selected by calculating the relative travel time in the Euclidean plane, considering the locations of the cardiac arrest patient and each ambulance. Furthermore, the MILP model is employed to rapidly identify optimal locations for drone ambulance stations out of 25 fire stations in Seoul, with a focus on 13,039 cases of cardiac arrests in 2022.

4 - Impact of Shift Structure on Fatigue among Emergency Department Service Workers

Chia-Chun Yang, University of Cincinnati, Cincinnati, OH, United States, Craig Froehle, Elizabeth Leenellett

It is well-established that worker fatigue can have deleterious consequences for work quality. However, less is known about how shift structure – an operational policy decision – influences worker fatigue. We collected primary data through a month-long fatigue-monitoring program we developed and implemented in an emergency department at a large, midwestern, academic hospital. Using a combination of nonparametric and econometrics models, we found that both the duration of a shift and its timing (i.e., when the shift starts) have important and differential impacts on workers' perceived physical and mental fatigue. In addition, we observed that time spent on either extra-clinical professional responsibilities or personal chores reduces the efficacy of a clinician's recovery time in terms of fatigue reduction. In contrast, we find that time spent with family and loved ones can significantly improve the effectiveness of fatigue recovery periods between work shifts. We conclude that scheduling policies that mitigate the physical and mental fatigue induced by shift work and allow for enhanced fatigue recovery between shifts should be preferred by organizations concerned with both employee well-being and consistently high service quality.

MC46

Summit - 438

Medical Decision Making

Invited Session

Health Applications Society

Chair: Narges Mohammadi, Imprial College Business School, London, United Kingdom

Co-Chair: Reza Skandari, Imperial College London, London, United Kingdom

1 - Sequential Updating Machine Learning Models for Diagnostic Testing

Kevin Smith, Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, United States, Erkin Otles, Siqian Shen, Brian Denton

Machine learning (ML) models are often used to recommend whether or not a patient should receive a diagnostic test. However, ML models are subject to *concept drift* whereby model performance degrades over time. One way to address concept drift is to update the ML model using recent data. However, in such cases the available data is subject to verification bias because results are only available for patients who received the diagnostic test. In this presentation, we discuss approaches for trading off the competing goals of model updating to improve referral decisions and avoiding unnecessary diagnostic testing for patients with a low risk of disease. We illustrate the approaches using a case study in the context of prostate cancer diagnosis.

2 - Unsupervised Learning Approaches to Detect Patterns in the Co-Occurrence of Diseases: Evidence from Medical Expenditure Panel Survey (Meps)

Ali Jafari, University of Massachusetts, Amherst, Amherst, MA, United States, Hari Balasubramanian

Understanding the relationships between diseases is one of the crucial aspects in healthcare and clinical practice. While certain associations between diseases, such as diabetes and hypertension (high blood pressure), are well established, others remain less obvious, and may remain undetected. In this study we use the Medical Expenditure Panel Survey (MEPS 2016 - 2019) to study associations between diseases. The MEPS provides comprehensive patient information, including specific medical conditions indicated by their corresponding International Classification of Diseases, Tenth Revision (ICD-10) codes. We structured a dataset from the MEPS records, facilitating access to individual disease profiles. Subsequently, we applied the Apriori algorithm to identify various disease combinations and association rules. Moreover, we implemented a novel unsupervised learning algorithm to perform clustering, offering deeper insights into the interrelations among diverse disease categories. Using these methodologies, we aim to enhance our understanding of disease relationships, potentially contributing to more effective healthcare strategies and interventions.

3 - Interpretable Mortality Simulation and Decision-Making Agent for Heart Transplantation

Yingtao Luo, Carnegie Mellon University, Pittsburgh, PA, United States, Rema Padman, Reza Skandari, Arman Kilic

Decisions regarding acceptance of a donor's heart for a patient on the transplant waitlist is currently based on evaluation by a committee of doctors who review many factors but are still quite ad-hoc. Given the volume and range of data collected by the United Network for Organ Sharing (UNOS) on waitlisted patients, donors, clinicians and organs, there is increasing interest among all stakeholders, particularly clinicians, in leveraging model-based approaches that learn from the vast amount of historical data on organ acceptance/rejection to optimize the heart transplantation outcomes and assist human decision making. In response to this need, we benchmark and propose machine learning models to provide accurate and interpretable predictions regarding transplant mortality at both the individual and population levels. We train our model with more than 30,000 patient records and 350 variables. This robust training has enabled the best model to achieve a C-statistic of 0.9, significantly better than previous benchmarks which only attained a C-statistic between 0.62 and 0.75 with 20 to 30 variables. This level of accuracy provides a reliable foundation upon which clinicians can base their decisions with a greater degree of confidence. We have also identified the key features that drive the predictions and verified the clinical significance of these results through close collaboration with the domain experts.

4 - Optimizing Colorectal Cancer Surveillance Policies Considering Uncertainty in Colonoscopy Outcomes

Mahboubeh Madadi, San Jose State University, San Jose, CA, United States, Jaleh Soltanpour, Amin Asadi

Colorectal and rectum cancer (CRC) is a significant cause of morbidity and mortality
worldwide. This work aims at optimizing colorectal cancer surveillance policies. Specifically, we propose a model to maximize a patient's quality adjusted life years (QALYs) through personalizing CRC surveillance strategy informed by patient's colonoscopy history. We propose a partially observable Markov decision process (POMDP) to model CRC dynamics. The proposed model incorporates the uncertainty in size and number of detected polyps in a colonoscopy test. We use data of over 47,000 colonoscopy reports (over 21,000 patients from 5 VA medical centers) to estimate the proposed POMDP parameters. Incorporating the detailed pre-cancerous states (i.e., size and number of detected polyps) results in a POMDP model with a large state space. We use different approaches such as approximate point-based value iteration and Reinforcement learning to find personalized optimal surveillance policies.

5 - A Machine Learning Model for Mortality Prediction of Paralytic Ileus Patients Admitted to ICU

Houshang Darabi, University of Illinois-Chicago, Chicago, IL, United States, Martha Razo

Paralytic Ileus (PI) patients in the Intensive Care Unit (ICU) are at a significant risk of death. Prediction of at-risk patients for mortality after 24 hours of admission of ICU PI patients is important to increase the life expectancy of PI patients. We propose a model called DLMP (Deep Learning Model for Mortality Prediction) that can predict the mortality event of the PI patients after 24 hours of admission to ICU. DLMP is a powerful deep learning model consisting of six total unique clinical lab items and two demographics as inputs to a Neural Network (NN) of only two neuron layers. Using the Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset of 1,017 ICU PI patients, the DLMP results in the best prediction performance with an AUC score of 0.889. DLMP framework significantly improves the prediction of mortality for PI patients compared to process mining and other machine learning models. The proposed DLMP has the potential to allow clinicians to create targeted interventions that reduce mortality for PI patients in an ICU setting.

MC47

Summit - 439

Analytics for Pharmaceutical Shortages

Invited Session

Health Applications Society

Chair: Daniel Kosmas, Northeastern University, Boston, MA, United States

Co-Chair: Noah Chicoine, Northeastern University, Boston, MA, United States

1 - When Should the FDA Inspect Pharmaceutical Manufacturing Facilities to Better Mitigate Drug Shortages?

Daniel Kosmas, Northeastern University, Boston, MA, United States, Ozlem Ergun

Drug shortages have been a persistent problem in American healthcare for decades, and the resulting lack of access to necessary drugs has been disastrous to patient health. A majority of these shortages were caused by quality issues related to problems in the manufacturing process. More frequent inspections can help reduce quality concerns, but deciding when to inspect is a complex problem; strict regulation enforcement can force low-profit facilities to close due to excessive maintenance costs, while lax enforcement allows for regulation violations to persist, both of which can cause drug shortages. We propose a novel POMDP model to assist inspectors in determining how long to wait between subsequent inspections that balancing the trade-off of recurring inspection costs, the benefits of allowing a facility to stay in operation, and the risks of allowing for disruptive events. We theoretically show that this problem can be reduced to only needing to consider whether or not to inspect immediately, which is independent of the time horizon. Our computational study highlights the importance of allocating more resources to high-risk facilities that produce drugs that highly impact public health. We additionally find that optimal inspection time is more sensitive to changes in the penalty from an unexpected disruptive event occurring the longer it has been since the last inspection.

2 - A Deep Reinforcement Learning Aided Inventory Control Approach for Managing Drug Shortages: Impact of Information Sharing and Adaptive Reward Shaping

Zohreh Raziei, Dematic, Plano, TX, United States, Ozlem Ergun

Pharmaceutical supply chains have long faced challenges in maintaining resilience, leading to drug shortages and inefficient inventory management. To address this issue, our study proposes a novel approach to inventory control that utilizes a multiagent deep reinforcement learning decision support system. This system models agents' behaviors and investigates the effects of partial and full information sharing during supply chain disruptions. We implement a reward-shaping strategy that helps increase stability during the training process and obtain more accurate policies under long and severe disruption scenarios. We proposed an automated weight adaptation strategy for the shaped reward, enabling the system to learn effective policies for maintaining supply chain resilience and mitigating the impact of disruptions on drug shortages.

3 - Assessing the Feasibility of an Intra-hospital Information Sharing System to Improve Drug Shortage Management

Noah Chicoine, Northeastern University, Boston, MA, United States, Jacqueline Griffin

During some drug shortages, hospital pharmacies receive little to no information regarding the cause and pervasiveness of the shortage. With such information, pharmacists could draw important inferences, such as future availability of the product and its alternatives, to guide their sourcing and ordering decisions. In this study, we develop a game-theoretic model of a supply chain system that aims to enhance hospitals' views of broader drug shortage market dynamics. In the model, hospitals anonymously share inventory and order information with each other during a drug shortage and make sourcing decisions based on that information. We test the feasibility of such an information sharing system and explore potential consequences, such as hoarding or external market consequences. The information-sharing system presented in this work could be incorporated into current inventory management software to help build resiliency in pharmaceutical supply chains. Hospitals, with the ability to make more informed inventory management and sourcing decisions, may be able to achieve more optimal care for their patients during drug shortages.

4 - Equitable Drug Distribution Under Panic Buying

Emily Tucker, Clemson University, Clemson, SC, United States, Marta Wosinska, Dan Li

When there is a signal that a drug shortage is imminent, health providers may rapidly begin to panic buy medications. As a response, wholesalers may impose order limits based on historical demand to spread available supply across providers. Yet, these approaches may not be sufficient to ensure that slower-responding health providers, e.g, less sophisticated systems, independent hospitals, or rural facilities, also receive medically-necessary drugs during shortages. In this project, we (i) study panic buying behaviors of health providers, including the characteristics of providers that receive drugs, and (ii) evaluate potential order limit policies to improve equity in drug distribution under different panic buying settings. Data is provided by a large drug wholesaler.

MC48

Summit - 440

Frontiers in Inventory Control: Asymptotic Analysis and Demand Learning

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Jinzhi Bu, The Hong Kong Polytechnic University, Hong Kong, N/A

1 - Asymptotic Optimality of Projected Inventory Level Policies in Lost Sales Inventory Systems

Joachim Arts, University of Luxembourg, Luxembourg City, Luxembourg, Poulad Moradi, Melvin Drent

We study the canonical periodic review lost-sales inventory system with positive leadtime and i.i.d. demand. We demonstrate that the relative value function, under the constant order policy and the long-run cost rate criterion, satisfies the Wiener-Hopf equation. We employ ladder processes associated with a random walk featuring independent and identically distributed (i.i.d.) increments, to obtain an explicit solution for the relative value function. Our analysis reveals that as lead time and penalty costs increase significantly, the value function tends to converge towards a quadratic form, irrespective of the demand distribution function, given that it possesses finite first and second moments. Furthermore, we show that under this regime, the projected inventory level policy outperforms the constant order policy.

2 - Asymptotic Optimality of Open-Loop Policies in Lost-Sales Inventory Models with Stochastic Lead TIMES

Xingyu Bai, University of Illinois at Urbana-Champaign, Urbana, IL, United States

In this paper, we consider the lost-sales inventory model in which the lead time is not only large but also random. Under the assumption that the placed orders cannot cross in time, we establish the asymptotic optimality of constant-order policies as the lead time increases for the model with divisible products. For the model with indivisible products, we propose a bracket policy, which alternates deterministically between two consecutive integer order quantities. By employing the concept of multimodularity, we prove that the bracket policy is asymptotically optimal. Our theoretical results can be extended to the models with order crossover, random supply functions, Markov-modulated demands and cyclic demands. Finally, we provide a comprehensive numerical study to demonstrate the good performance of the proposed open-loop policies, and derive further managerial insights.

3 - Managing Perishable Inventory Systems with Positive Lead TIMES: Inventory Position vs. Projected Inventory Level

Huanyu Yin, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Jinzhi Bu, Xiting Gong

We consider periodic-review perishable inventory systems with a fixed product lifetime and positive replenishment lead times. Demands are satisfied by on-hand inventories of different ages following a general issuance policy. Unsatisfied demand is either backlogged or lost. The objective is to minimize the long-run average holding, penalty, and outdating costs. The optimal replenishment policy for these systems is notoriously complex and intractable due to the curse of dimensionality. Previous studies have shown that the class of base-stock policies, which maintains a constant inventory position in each period, performs poorly when the unit penalty cost is large. In this paper, we propose a more appealing class of projected inventory level (PIL) policies that maintains a constant expected on-hand inventory level at the time of each order arrival. For both backlogging and lost-sales systems under a general issuance policy, we prove that the best PIL policy is asymptotically optimal with large unit penalty costs under a large class of demand distributions. For the backlogging system, we further prove that its optimality gap decays to zero exponentially fast in the product lifetime and in the demand population size under a large class of issuance policies. Our numerical results show that the best PIL policy performs close to optimal and much better than the best base-stock policy.

4 - Bayesian Inventory Management with High-Dimensional Data

Weizhou Zhang, National University of Singapore, Singapore, Singapore, Hanzhang Qin, Jussi Keppo

We study Bayesian inventory control problems where the parameters of the demand distribution are known to belong to the exponential family a priori, but the parameters need to be learned from demand samples. We consider a finite-horizon perishable inventory control problem setting with unobserved lost sales, thus the sales data are censored by the inventory level. Besides estimating the demand parameter from the posterior belief, we derive the maximum likelihood estimation (MLE) approach to estimate the demand parameter using the right-censored sales data. Indeed, we show that the posterior mean can be recovered from the MLE expression when we have a sufficiently large number of data samples. Meanwhile, we develop a data-driven scheme to predict the stock-out probability (i.e., the predicted service level) at each retailer given any base-stock inventory control policy -- we show the probability can be uniquely identified from the data. Moreover, we extend the model, algorithm, and analysis to a joint price and inventory Bayesian model which is an extension of the existing single-parameter Bayesian inventory model. We conducted the numerical experiment against synthetic data for single and two-dimensional Weibull distribution data. An approximate inference scheme is applied when the dimensionality of the latent variable space is high and no closed-form analytical solution exists.

5 - Partial Backorder Inventory System: Asymptotic Optimality and Demand Learning

Zhaoxuan Wei, Institute of Operations Research and Analytics, NUS, Singapore, Singapore, Andrew Lim, Hanqin Zhang

We develop a stochastic inventory system which accounts for the limited patience of backlogged customers. While limited patience is a feature that is closer to the nature of unmet demand, our model also unifies the classic backlogging and lost-sales inventory systems which are special cases of the one we propose. We establish the uniform (asymptotic) optimality of the base-stock policy when both demand and patience distributions are known. When the backlogged demands become unobservable, we introduce a novel policy family that operates without backlogged demands information, and prove that it can approach the cost efficiency of the optimal policy in the system when the demand and patience distributions are known. Finally, we consider an online inventory control problem in which backlogged demand is unobservable and demand and patience distributions are also not known, and develop a UCB-type algorithm that yields a near-optimal policy.

The regret bounds given by the algorithm are provably tight within the planning horizon, and are comparable to the state-of-the-art results in the literature, even in the face of partial and biased observations and weaker system ergodicity.

MC49

Summit - 441

Social Impact of Operations Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Zhihan Wang, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

1 - Maternal Health Interventions with Combinatorial Restless Bandits

Lily Xu, University of Oxford, Oxford, United Kingdom, Bryan Wilder, Elias Khalil, Milind Tambe

Restless bandits are a useful model for many real-world problems, whose deployments include for maternal healthcare interventions in India. However, existing approaches have been restricted to simplified action settings that allow for tractable solutions. Unfortunately, public health interventions often have complex combinatorial constraints — such as path constraints or scheduling constraints — that cannot be addressed by existing solvers.

Here, we consider CoRMAB, a broader class of problems with combinatorial actions that cannot be decoupled across the arms of the restless bandit, requiring direct solving over the joint action space. Leveraging recent advances in embedding trained machine learning models into optimization problems, we propose SEQUOIA, an algorithm for training a reinforcement learning agent to find a good policy that considers long-term reward by embedding a Q-network directly into a mixed-integer program to select one of exponentially many combinatorial actions in each timestep. SEQUOIA builds on the widely used deep Q-network framework but also benefits from a set of enhancements that accelerate training by exploiting the structure of CoRMAB. We empirically validate this approach on novel restless bandit problems with combinatorial constraints, including from maternal healthcare in India.

2 - Disruptions, Resilience and the Performance Gender Gap in Small Firms

Amrita Kundu, McDonough School of Business, Georgetown University, Washington, DC, United States, Kamalini Ramdas, Stephen J. Anderson

Why a large gender gap in firm performance persists – especially for small firms in developing countries – is not well understood. In this paper, we study the operating environment of such firms to assess to what extent firm-specific disruptions contribute to the performance gender gap. Further, we examine to what extent building resilience can help women-led firms improve performance. In a sample of 646 small firms in Kampala, Uganda, we find evidence of a sizeable gender gap in the downside losses from firm-specific disruptions, which reduce managerial capacity and resource availability. On average, business disruptions decrease short-term sales and sales growth of small women-led firms by an additional 11.6% (p=0.013) and 15.2 percentage points (p=0.084), respectively, relative to those of men-led firms. In the longer term, disruptions alone explain 8% of the profit gender gap. Building resilience helps women-led firms buffer against the negative impact of disruptions. In the absence of resilience, women-led firms that face managerial disruptions see a decline in sales of 13.4% (p = 0.069) and those that face operational disruptions lose 48% in sales (p = 0.001) in a six-month period. In contrast, women-led firms with high resilience fully overcome these disruption-related losses. In our sample, women-led firms are 9% less likely than men-led firms to have high relational resilience (p = 0.000) and 5% less likely to have high resource resilience (p = 0.002). Our results indicate that substantial gains could accrue from instituting resilience training for women-owners of small firms.

3 - Impact of in-Class Smartphone Bans on Student Outcomes

Andreas Bjerre-Nielsen, Univeristy of Copenhagen, Copenhagen, Denmark

We evaluate a novel large-scale RCT on in-class smartphone bans. We investigate its impact on students' academic outcomes in terms of grades obtained as well as class attendance and student wellbeing.

4 - Optimization Meets Participation: Iterative Zone Generation for School Assignment

Irene Lo, Stanford University, Stanford, CA, United States

In U.S. public schools, geographic boundaries significantly influence student-school assignments. However, such boundaries often replicate residential segregation patterns in schools. To address this, the San Francisco Unified School District (SFUSD) adopted guidelines for multi-school zones, balancing neighborhood assignment and citywide choice to mitigate segregation. Optimizing these zones is computationally and socially complex due to their direct and indirect impacts on policy goals. We propose an iterative framework that promotes stakeholder participation in optimizing multi-school zones. Our approach alternates between a consumer search process using interactive tools for constraint identification and selection, and computational methods generating solutions satisfying those constraints. We develop novel computational approaches for the multi-school zone problem using both optimization and sampling-based methods. Comparing our approaches to benchmarks from the literature, we find that our multilevel optimization approach consistently produces better zone-level statistics for larger zones, while our recursive optimization-based approach excels in generating numerous small zones. Our framework supported SFUSD's development of a zone-based policy and is being used to facilitate stakeholder participation in determining the final zones.

5 - Closer Childcare Better Career: An Empirical Study of Childcare Availability and US Knowledge Workers

Damian Beil, University of Michigan, Ann Arbor, MI, United States, Zhihan Wang, Jun Li

The scarcity of quality childcare services has long obstructed the career progression of working parents in the US. Leveraging large-scale national job profile data from 1997 to 2019, we investigate how the career development of employees in a variety of industries is influenced by the availability and location of childcare services nearby their working sites.

MC50

Summit - 442

Empirical Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: David Rea, Lehigh University College of Business, Bethlehem, PA, United States

1 - A Framework to Study Health Inequities: A Case Study of Breast Cancer Survival

Margret Bjarnadottir, University of Maryland, College Park, MD, United States, Ritu Agarwal, Wedad Elmaghraby, Shana Ntiri, Lindy Rosal, Nawar Shara

We present a health inequity framework that can summarize health inequities and guide action toward eliminating them, by highlighting drivers of inequity and their relative importance. We present a novel approach to linking unadjusted inequity to adjusted inequity highlighting both key quantifiable drivers and the degree to which the inequity is a results of structural and systemic racism, or other or unmeasurable or omitted factors. We offer the framework as a decision support tool for health systems to study their own outcomes and drive a deeper understanding of health inequity outcomes.

2 - Designed for Variability: An Empirical Study of a Dedicated Observation Unit

Temidayo Adepoju, Rutgers Business School, Newark, NJ, United States, Anita Carson, Cherisse Carlo, Chris Manasseh

Safety-net hospitals with focused units have to provide care for patients who do not fit the focused unit's parameters. We study how a dedicated observation unit transforms into a "variability-absorbing" unit to treat patients with comorbidities and more complex conditions. Studies on observation units have largely been based on the unit treating patients with no comorbidities. Higher-rate of reimbursement for patients who stay more than two midnights on inpatient units have triggered hospital managers to find capacity and resources to care for the patients who stay less than two midnights yet with comorbidities and complex conditions. Transforming the DOU into a "variability-absorbing unit" to admit these patients keeps the inpatient unit "focused" on patients who stay more than two midnights. Managerial intervention, such as external collaborative coordination (ECC) improves performance of the DOU. To study this empirically, we use data from Delta Hospital to quantify the effect of the DOU transformation. We find a decrease in ED boarding time and total per visit average reimbursement for the two midnights or more patients admitted into the inpatient unit. Next, we find that the DOU transforming into a "variability-absorbing unit" worsened length of stay (LOS) compared to its prior performance. However, through the implementation of ECC, we demonstrate that the DOU has the potential to perform efficiently with LOS eventually decreasing. We conclude that the DOU can serve as a strategy for hospital managers seeking to provide a path forward for creating focused units in hospitals.

3 - Learning from Quality Signal to Improve Fairness and Accuracy: Evidence from the Hospital Readmissions Reduction Program

Mohamad Soltani, University of Alberta, Edmonton, AB, Canada, Bob Batt, Hessam Bavafa

Policymakers often employ quality indicators to evaluate the performance of an organization and to provide incentives for quality improvements. A potential concern, however, is that a specific quality indicator may be unfair to agents facing different circumstances. We explore the effects of the Hospital Readmissions Reduction Program on 30-day readmissions in over 2,000 US hospitals. While the policy does not differentiate between readmission reductions within the 30-day time window, our findings show that the extent of improvement depends on the timing of readmission. We attribute this variation to the time-dependent costs associated with reducing readmissions. This finding, combined with the inconsistent readmission distributions across hospitals, shows that HRRP is unfair to safety-net hospitals that treat low-income patients with limited access to care.

4 - Time Vs Testing: Physician Information Gathering Under Variable Conditions

David Rea, Lehigh University, Bethlehem, PA, United States, David Anderson, Margret Bjarnadottir

Using a comprehensive data set from a large emergency department, we provide empirical evidence that physician's gather information in different ways depending how much time they have. Their behavior at the beginning and end of shift depends on the initial and terminal workload conditions. This has implications for patient care and hospital flow.

5 - Responsible Clinician-AI-Collaboration Framework to Build Trustworthiness AI in Health Systems.

Elham Nasarian, Virginia Tech, Blacksburg, VA, United States, Kwok-Leung Tsui

This systematic review investigates the use of interpretable machine learning (IML) and explainable artificial intelligence (XAI) in healthcare, focusing on clinical decision support systems (CDSS). AI technologies like medical sensors, health trackers, telemedicine, mobile health (mHealth), large language models (LLMs), and digital care twins (DCTs) significantly influence clinical decision-making. However, the complexity of medical decisions requires that AI-generated results be interpretable and explainable to clinicians, as a lack of interpretability can lead to mistrust and reluctance to use AI.

The study aims to examine the processes, methods, applications, and challenges of IML and XAI in digital health interventions (DHIs). It categorizes the IML process into pre-processing, interpretable modeling, and post-processing phases to foster a comprehensive understanding of robust interpretability in CDSS. The goal is to enhance communication between AI systems and clinicians.

Using the PRISMA guideline and the PICO method, the authors searched databases like PubMed, Scopus, and Web of Science. After screening and quality assessment, 74 publications were selected for analysis, including reviews and experimental studies.

The paper highlights the need for an interpretability framework in CDSS across three levels and explores XAI applications within each tier. It discusses tools for evaluating XAI in health systems and presents a roadmap for implementing XAI in clinical settings. The conclusion emphasizes the importance of XAI models and acknowledges their limitations, suggesting future research directions to bridge existing gaps.

MC51

Summit - 443

Supply Chain Finance and Risk Management

Invited Session

MSOM: iForm

Chair: Gangshu Cai, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Tingyi Yu, Shanghai Jiao Tong University, Shanghai, N/A

1 - Managerial Incentives, Partial Ownership, and Channel Decentralization

Bangdong Zhi, University of Bristol, Bristol, United Kingdom, Jiong Sun

This study proposes a simple mechanism in which passive ownership in the downstream firm of a vertical relationship fosters economic incentives for downward decentralization. This mechanism works by motivating intrafirm managerial efforts aimed at enhancing productivity. The enhanced productivity results in higher profits for the firm when it sells its product indirectly compared to selling it directly. Interestingly, this scenario happens when the managerial effort cost falls within an intermediate range—not too low, yet not too high. Furthermore, this mechanism remains effective in incentivizing downward decentralization even in the presence of competition, albeit by strategically demotivating managerial efforts as a means to soften competition. Finally, we demonstrate the effectiveness of this mechanism in motivating upward decentralization provided that the firm retains channel power.

2 - A Decision Support System for Social Media Online Harassment Using Blockchain

Edwin (Yuchen) Wang, University of Massachusetts Boston, Boston, MA, United States, Nesreen El-Rayes, Adeela Gulzari

Online harassment has become an increasingly severe issue, particularly impacting minority communities. According to the Anti-Defamation League (2022), 35% of Americans have experienced harassment that specifically targets their ethnic, religious, or sexual identities. While blockchain technology (BCT) has been extensively applied in commercial settings, this paper focuses on its potential to enhance information management for decision making to combat online harassment. Existing literature on BCT proposes frameworks with comprehensive design features; however, these often exhibit overlapping functionalities. The proposed BCT framework allows users to maintain anonymity yet undergoes a singular verification process to ensure accountable behavior without compromising privacy. Furthermore, the framework utilizes smart contracts to automatically enforce community standards. Users consent to these regulations upon registration, with breaches triggering automatic penalties. This is facilitated by consensus mechanisms that allow the community to confirm harassment reports. Additionally, the framework introduces a blockchain-based reputation system, wherein positive interactions boost user reputation, whereas verified harassment incidents result in penalties. From a decision analysis perspective, this paper integrates a systematic evaluation of strategic choices within the blockchain model, optimizing the balance between anonymity and accountability.

MC52

Summit - 444

Information Provision and Platforms

Invited Session

MSOM: Service Operations

Chair: Wenchang Zhang, Indiana University, Bloomington, IN, United States

1 - Optimal Information Provision and Competition among On-Demand Service Platforms

Wenchang Zhang, Indiana University, Bloomington, IN, United States

This research examines the transformative impact of on-demand service platforms on service delivery and consumption, emphasizing the strategic role of information policy design in managing platform dynamics, service provider decision-making, and consumer behavior. Through a model analyzing interactions across an infinite discrete-time horizon, the study investigates how platforms use information provision and commission policies to influence the market, particularly focusing on service quality. Analyses in both monopoly and competitive scenarios demonstrate that optimal information policies can simplify platform strategies and enhance competitive positioning by aligning service offerings with consumer expectations. In monopolistic markets, platforms can optimize their strategies using convex quadratic programming, while in competitive markets, platforms engage in a Stackelberg game to effectively manage information, thus optimizing provider participation and maximizing revenue. The findings highlight information policy as a crucial lever for platforms to navigate market complexities and secure a competitive advantage.

2 - Information Design of a Delegated Search

Yangge Xiao, Institute of Operations Research and Analytics , National University of Singapore, Singapore, Singapore, Zhenyu Hu, Shouqiang Wang

A principal delegates a sequential search with finite search opportunities to an agent, who bears the search cost and can terminate the search. The termination payoff is split between them according to a pre-specified proportion. Novel in our setting, only the principal can evaluate the search outcomes and design a policy to decide whether to continue the search and, if so, what information to provide to the agent after each search. Formulating the principal's problem by dynamic information design, we obtain a complete analytical characterization of her optimal information policy featuring a sequence of deterministic acceptance standards. The agent is recommended and voluntarily willing to continue the search if and only if the current termination payoff fails to meet the corresponding standard. For non-recallable search, the acceptance standards are descending and informative, underscoring the importance of timely feedback to the agent for non-recallable searches such as talent recruiting or housing hunting. In contrast, for recallable search, the optimal thresholds can become completely uninformative before a cutoff opportunity, where the agent searches in the dark. After the cutoff opportunity, the principal sets a sequence of descending acceptance standards by equating the agent's marginal cost and return from one additional search. For innovation-driven searches like pharmaceutical R&D, where search outcomes accumulate, our results suggest that search incentives are best motivated by an idealistic vision followed by realistic compromises. Overall, our findings show how costless information can be used as an incentive device in the absence of costly pecuniary instruments.

3 - Platform Endorsement and Customer Search: Evidence from Airbnb

Fanyin Zheng, Imperial College London, London, United Kingdom, Omar Besbes, Yuri Fonseca, Zhenyu Lai, Ralph Mastromonaco

We study the impact of platform endorsement on customer's search and purchase decisions on large two-sided service platforms. We model customer's behavior using a sequential search model and estimate the model using data from a large hospitality platform.

4 - Diverse Assortments in Online Recommendations

Mahsa Hosseini, University of Windsor, Windsor, ON, Canada, Opher Baron, Shreyas Sekar, Azarakhsh Malekian

This paper explores the business benefits of incorporating diverse content into recommendation systems, particularly in online marketplaces and platforms that rely on ranking algorithms. While such systems enhance short-term user engagement by promoting popular content, they inadvertently risk sacrificing diversity in recommendations, potentially jeopardizing the long-term health of the marketplace. This paper studies the interplay between consumer engagement and the phenomenon of "satiation" - a concept that examines how past consumption patterns influence consumers' future valuation of goods or services. Our research finds conditions under which a more diverse recommendation policy is optimal. Our findings show that optimal policy favors diversity during earlier periods. However, as time progresses, the optimal policy gravitates towards consistently offering products with the highest perceived value. This implies that a more diverse recommendation set is more desirable over extended time frames. Furthermore, we prove a satiation threshold policy, offering insights into determining optimal recommendations based on varying satiation levels at different periods. By incorporating diversity into our satiation model, we show that our optimal recommendation strategy not only promotes diversity and equity for consumers but also improves platform profitability.

MC53

Summit - 445

Advances in Sustainable Supply Chains: Theory and Practice

Invited Session

MSOM: Supply Chain

Chair: Li Chen, Cornell University, Ithaca, NY, United States

Co-Chair: Shiqing Yao, Monash University, Melbourne, Australia

1 - Converting Counterfeiters in Emerging Markets to Authorized Suppliers

Burak Kazaz, Syracuse University, Syracuse, NY, United States, Liling LU, Xin Fang, Yini Gao

In recent years, “super fakes,” counterfeits of high quality, have become popular. Super fake manufacturers’ capability to produce high-quality products inspires a new anti-counterfeiting measure: converting counterfeiters to authorized suppliers. To study the effectiveness of this anti-counterfeiting measure, our paper employs a game-theoretic model to examine the interactions between a brand-name firm with its home supplier and a counterfeiter who produces high-quality counterfeits but can be potentially converted to an authorized overseas supplier. Our results show that a counterfeiter can be converted to an authorized overseas supplier by dual sourcing or single sourcing, if the difference in production costs between two suppliers or the penalty from law enforcement is high. Otherwise, the brand-name firm may have no incentive to convert the counterfeiter due to limited cost savings and due to the high wholesale price needed. However, the overseas supplier may still decide to enter the market to sell the counterfeit if the penalty from law enforcement is low. With a smaller difference in the perception of quality from the brand-name and counterfeit, we demonstrate that it is easier for the brand-name firm to deter the overseas supplier's counterfeiting by single sourcing from overseas supplier than by dual sourcing. We also examine the impact of this anti-counterfeiting measure on consumer and social surplus, and find that converting a counterfeiter to an authorized supplier may hurt consumer surplus and does not always improve social surplus.

2 - Consumer Status Signaling, Wealth Inequality and Non-deceptive Counterfeits

Shiqing Yao, Monash University, Melbourne, Australia, Li Chen, Zhen Lian

The rising of ecommerce has made it easy for consumers to search and purchase cheap non-deceptive counterfeits to send a “deceptive” status signal, posing a new challenge to the luxury (status product) industry. Motivated by these industry dynamics, we consider a market entry deterrence game between an incumbent status product firm and a non-deceptive counterfeiter who attempts to enter the market. Our analysis yields three main insights. First, without counterfeits, the firm is strictly better off from the heightened motive of consumer status signaling; however, such benefit would be neutralized by the potential counterfeiter entry. Second, the presence of counterfeits lowers the firm’s profit, but may induce the firm to raise its price. Third, the presence of counterfeits has mixed effects on social welfare; only when the wealth inequality among consumers is moderate, will social welfare increase. The findings from our analysis offer some initial guidelines to managers of luxury brands and e-commerce platforms in addressing the nondeceptive counterfeit problem.

3 - Impacts and Implications of Transformative Adaptation in Food Systems in Response to Severe Droughts

Bingyan Dai, Cornell University, Ithaca, NY, United States, Miguel Gomez, Harry Kaiser

Drought is one of the most significant drivers of food production shocks. It can impact various aspects of food systems, from crop yields, food distribution and market access to consumption. Transformative adaptation provides potential solutions by fundamentally altering agricultural systems in response to actual or expected climate shocks. However, there is limited understanding of how such transformation impacts economic and environmental performances of food systems, particularly in fresh vegetable supply chains. This research examines the impact of transformative adaptation on food systems in response to increased drought risks, focusing on fresh vegetables in the United States. Specifically, we utilize a production-transportation optimization model to examine the fresh broccoli supply chain in the eastern U.S. market, a centralized supply system vulnerable to drought-induced disruptions in California. We find that while increasing drought severity could substantially decrease broccoli shipments from its major supplier, California, the reallocation of production from California to the East Coast in response would modestly decrease total supply-chain costs (by 1.4% in the worst-case scenario) in the eastern market. Our results also identify optimal production locations, production volumes and product flows during two drought scenarios. The method and findings contribute insights applicable to other vulnerable fresh vegetable supply chains, and inform policymakers and food industry stakeholders on effective adaptation strategies for building resilient food systems amidst drought risks.

4 - Ultra-Fresh Fashion: Creating Demand with Freshness and Agility

Hau Lee, Stanford University, Stanford, CA, United States, Li Chen, Shiqing Yao

Using agile supply chains, fast fashion companies have been viewed as best-practice examples in industries. Prior research has focused on how agility can equip such companies with strong “sense and respond” capabilities to identify and fulfill unpredictable customer demands. There is another powerful dimension of agility—the ability to create new products frequently—that has enabled recent market success of companies such as Shein. We seek to model this dimension of agility as an operational strategy for demand creation. Our model enables us to explore the impacts through the lens of profit to the firm, consumer surplus, and environmental performances.

MC54

Summit - 446

Operational Insights into Innovation, Economics, and Digitalization

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Xiaojin Liu, Virginia Commonwealth University, Richmond, VA, United States

Co-Chair: Pankaj Kumar, Virginia Tech, Blacksburg, VA, United States

1 - Operations Safety Innovations and Applications: the Entrepreneurial Journey of Immune Material Imu+

Chris K. Y. Lo, The Hong Kong Polytechnic University, PolyU Business School (PBS), Hong Kong, Hong Kong, Chi Wai Kan

The ongoing challenges of infectious diseases, highlighted during the COVID-19 pandemic, emphasize the crucial need for innovative safety solutions. This presentation outlines the entrepreneur journey of Immune Material Limited (IMU+). This venture was co-founded by Prof. Chris K. Y. Lo, who conducted years of research in operational safety risks, health and safety violations, safety management systems, and supply chain risks into the creation of advanced antimicrobial material technologies. The technology can potentially control health hazards in health care and service operations. Utilizing a risk-based innovation framework, Prof. Lo pioneered these technologies to address critical public health safety needs across various high-contact environments in Hong Kong and China, creating values from identifying and mitigating operational risks.

IMU+'s antimicrobial materials have been adopted in over 300 Hong Kong facilities, including public hospitals, hotels, schools, and shopping malls. It can dramatically reduce pathogen transmission and enhance operational safety with less manpower. This technology has successfully secured substantial private and public funding and market adoption, demonstrating its effectiveness and broad applicability in healthcare, schools, and other public infrastructures. This case study showcases how the strategic application of a risk-based innovation framework can improve operations safety and efficiency and set new operations standards for public health and safety.

2 - Mandatory Environmental Disclosure and Environmental Performance

Jie Lian, The University of Oklahoma, Norman, OK, United States, Yan Dong, Keith Skowronski, Sining Song

Regulators frequently use mandatory environmental disclosure regulations to encourage corporate environmental improvements. This research explores the broader implications of such mandates, examining how they may lead firms to reallocate resources and make technological trade-off decisions across different environmental domains. Our findings illuminate the multidimensional impacts of these regulations, offering insights into their effectiveness and potential unintended consequences.

3 - Routines, Disruptions and Organizational Adaptation

Rob Glew, McGill University (Desautels), Montréal, QC, Canada, Jeremy Hutchison-Krupat

It is well-established that routines and norms help to drive organizational efficiency. It is also well-established that organizations need to adapt. The need to adapt becomes even more salient when a disruption has occurred. However, disruptions do not necessarily require organizational norms and routines to change; organizations that abandon valuable routines unnecessarily will underperform. In this paper we build a model to study how organizations adapt to disruptions and characterize an organization’s propensity to adapt under different conditions and the likelihood that an organization adapts unecessarily.

4 - How Do Managerial and Operational Governance Impact Firm Innovation?

Tianxiu Zhang, University of Science and Technology of China, Hefei, China, People's Republic of, Zhijian Cui, Xiaojin Liu

Corporate governance can be divided into two distinct types: managerial governance and operational governance. Scholars of managerial governance focus on addressing agency problems and aim to regulate the decision-making process of senior managers and control the allocation of a firm's resources. Conversely, scholars of operational governance concentrate on how firms use process management techniques (such as TQM, Six Sigma, and other quality controls) to regulate operational routines and strategy implementation, thereby improving operational efficiency and innovation. The effectiveness of governance relies on the integration and coordination of these different levels. In this study, we empirically investigate whether and how these two types of governance affect firms' innovation performance, considering both exploitative and exploratory innovation. To do this, we develop proxies for corporate governance and operational governance using a panel sample of Chinese listed companies from 2007 to 2020, leveraging governance indices and China Quality Award acceptances and awards. Our analysis incorporates the xgboost machine learning model as a part of our empirical strategy. The results suggest that both managerial and operational governance are positively associated with exploitative innovation; however, operational governance has a detrimental effect on exploratory innovation. Our findings also highlight the significant substitution and/or complementary effects between managerial and operational governance when it comes to different types of innovation.

5 - An Empirical Study of Organic Product Spillover in Assortment

Zhihao Zhang, University of Missouri - Kansas City, Kansas City, MO, United States, Yan Dong, Sriram Venkataraman, Jayanth Jayaram

We study how introducing and increasing organic products affect conventional product assortment and revenue at retail stores. We address this research question by using proprietary scanner data with respect to organic and conventional yogurt sales spanning four years.

MC55

Summit - 447

Analytics for Societal Impact

Invited Session

Public Sector OR

Chair: Shima Azizi, INFICON, Westbury, NY, United States

1 - Optimizing Foster Care Visitation Scheduling

Shima Azizi, INFICON, Westbury, NY, United States, Andrew Trapp, Erhun Kundakcioglu, Caroline Johnston

Child welfare agencies charged with ensuring foster children regularly visit biological parents can be challenged due to fixed workforce levels and fluctuating caseloads. We introduce the Foster Care Visitation Scheduling Problem to assign, schedule, and route workers to foster children. We develop a two-phased network-based optimization approach that 1) preprocesses and pre-computes a time-space network structure, and 2) solves a novel large-scale integer optimization problem over this network. We incorporate our approach into an interface, assists foster care organizations to better operationalize their resources, and improves the consistency of visits and so quality of life for foster children. We discuss computational experiments on a variety of instances inspired by real data from New York State that reveal encouraging computational performance.

2 - Disaster Relief Inventory Simulation: Managing Resources in Humanitarian Camps

Erhun Kundakcioglu, Ozyegin University, Istanbul, Turkey, Cem Yarkin Yildiz

Disasters often displace populations into temporary settlements, creating complex challenges in post-disaster inventory management. While existing research primarily focuses on isolated aspects, such as funding variability, there remains a significant gap in adopting a holistic approach. This lack of comprehensive understanding limits the validation and applicability of proposed policies to real-world scenarios.

To address these limitations, this study introduces an extensive, event-based, open-source simulation framework tailored for the unique aspects of humanitarian inventory management. The simulation framework models the problem from the perspective of a central decision-maker, responsible for decisions on replenishment and transshipment, aiming to satisfy the demands of multi-camps for multi-items. It integrates a full spectrum of dynamics—including the perishability of items, demand, migration, funding with earmarked and in-kind options, and supply disruptions—without reliance on Markovian assumptions.

All events are generated based on stochastic properties with any distribution and parameters. The framework's engine manages event dependencies, ensuring that any state change triggers necessary events to regenerate to satisfy non-Markovian properties. Users can configure their settings and implement their policies with ease, and detailed performance indicators for simulations are provided. This enables realistic simulations of complex humanitarian scenarios and facilitates thorough policy evaluations.

The framework not only advances theoretical research but also equips practitioners with a robust tool for simulating diverse scenarios, thereby enhancing practical decision-making in disaster response.

3 - Optimizing Surplus Medicine Donations: Corporate Social Responsibility and Supply Chain Coordination

Telesilla Kotsi, The Ohio State University, Columbus, OH, United States, Soh Hyun Chu, James Hill

Pharmaceutical firms are essential in improving health through medicine donation programs that help people in the U.S. and around the world. This study looks at donations of prescription drugs for chronic conditions in the U.S., where the high cost of medications can be challenging for many. Brand-name prescription drugs cost an average of $334 per month, compared to $58 for generics, making affordability a problem for many, especially seniors and the uninsured. We examine the "push" supply chain model used for surplus medicine donations, focusing on pharmaceutical firms since they determine the quantities of donated drugs that flow through the rest of the supply chain. Our extensive discussions with many firms revealed that effective collaboration between Corporate Social Responsibility and Supply Chain teams is crucial for managing these donations and ensuring that medicines with limited shelf lives (6-18 months) reach those in need. We use a game-theoretical model and lab experiments to understand how supply chain managers make decisions and identify incentives that align their choices with the firm's social sustainability goals. This method explains how personal traits influence donation decisions and offers insights into improving supply chain strategies for both economic and social sustainability. The study emphasizes the importance of coordinated efforts to address the medication needs of vulnerable populations and ensure access to medicines for the most vulnerable.

4 - Optimization Models for 15 Minute Cities

Farima Salamian, University of Iowa, Iowa City, IA, United States, Ann Campbell

The 15 Minute City(15MC) framework aims to enhance urban livability by ensuring access to daily needs for residents, such as schools, parks, grocery stores, and medical offices, within a fifteen-minute walk. The 15MC concept promotes sustainability and improves quality of life. Our study advances the implementation of the 15MC concept by modeling the selection of proposed locations for various key amenities for city planners. Our research employs mixed integer linear programming to optimize the location of new amenities, considering different objective functions that reflect maximizing accessibility. We test our models using data from Iowa City, including walking distances and the locations of current amenities. Our results provide insights regarding the implications of different objective functions and suggest several directions for future research.

MC56

Summit - 448

Writing Strategies

Panel Session

Minority Issues Forum

Co-Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, 80401, United States

1 - Moderator Panelist

2 - Panelist

Candace Yano, UC Berkeley, Berkeley, CA, United States

x

3 - Panelist

Jamol Pender, Cornell University, Ithaca, NY, United States

x

4 - Panelist

Erick Jones, University of Texas at Arlington, Arlington, TX, United States

x

5 - Panelist

Oluwaseun Ogunmodede, The MITRE Corporation, McLean, VA, United States

x

MC57

Summit - Terrace Suite 1

Empirical Research in Healthcare Analytics

Invited Session

Health Applications Society

Chair: Jong Myeong Lim, Miami Herbert Business School, Coral Gables, FL, United States

1 - The Unintended Costs of Congestion Mitigation Through Financial Barriers: Empirical Evidence from Online Health Communities

Jie Chen, Arizona State University, Tempe, AZ, United States, Kan Xu, Tongxin Zhou

While information technology fosters greater and equal accesses to consumers by creating a decentralized market, it also raises challenges due to market congestion where high screening costs suppliers have to pay to enhance matching quality. This paper examines the use of financial barriers to address market congestion in Online Health Communities (OHCs), a setting where congestion issues have especially significant implications for societal welfare and health equity. By leveraging a natural experiment within the POZ community, we provide causal evidence of both the intended and unintended effects of financial barriers implementation on user participation. Our findings show that while financial barriers reduced inquiries by approximately 17.5 percentage points in the treated forum, they also led to significant unintended consequences. Specifically, there was a 15.8 percentage point decrease in the likelihood of users asking any questions in the treated forum, accompanied by a 5.8 percentage point drop in demand and a 6.1 percentage point decrease in posting likelihood in untreated forums. Further analysis reveals that these negative spillovers are consistent across users with varying tenures and usage frequencies. Additionally, results for users' behavioral responses after the policy announcement show that users form questions with more effort and negative emotions. Our study contributes to the congestion mitigation literature by highlighting crucial negative spillovers associated with financial barriers in OHCs. Contrary to existing research suggesting that minimal costs can improve matching quality, our findings underscore the complexities and potential drawbacks of implementing financial barriers, challenging the notion that they universally enhance platform efficiency.

2 - How Vertical Integration Effect the Efficiency of the Hospitals in the U.S

Jingyun Su, Miami Herbert Business School, Coral Gables, FL, United States, Jong Myeong Lim

This study investigates the effects of vertical integration on the efficiency of hospitals within the United States. Using Data Envelopment Analysis (DEA), we compute efficiency scores for a diverse set of hospitals to determine their operational effectiveness relative to the best-performing entities. Further, we employ regression analysis to explore the relationship between vertical integration and efficiency scores, controlling for various hospital characteristics such as size, and location. Preliminary findings suggest that vertical integration may contribute to enhanced operational efficiency, potentially driven by improved resource allocation and streamlined management practices. This analysis not only sheds light on the operational dynamics of vertically integrated hospitals but also informs policy discussions on healthcare management and organizational structure. The implications of these findings for healthcare policy and management are significant, suggesting pathways for enhancing hospital performance through structural adjustments.

3 - How Recreational Cannabis Legalization Impacts Hospital Operations

Max Yakovlev, Kellogg School of Management, Northwestern University, Chicago, IL, United States, Maria Ibanez

Cannabis is the most common federally illegal drug in the US. The federal stance contrasts with the growing trend of recreational cannabis legalization (RCL) across states. A vital implementation question is how legalization affects hospitals. We propose that as cannabis spreads throughout society, its effects permeate healthcare systems, particularly inpatient wards. Based on medical research, patients using or exposed to cannabis will suffer from challenging symptoms, including aggravated chronic conditions, psychosis, agitation, and intractable vomiting. These peculiar symptoms will increase workload complexity, restrict the staff time and hospital capital resources available to other patients, and cause disturbance. To unpack these effects, we study how RCL influences hospital operations across three dimensions: throughput times (length of stay), experiential quality of care, and medical providers' behaviors. Using a difference-in-differences approach, we find that hospitals speed up care, with legalization shortening inpatient lengths of stay by 2.6%. We attribute this phenomenon to cannabis generating workload for hospital staff that reduces the resources (including staff time) available to other patients. Second, we find adverse effects on experiential quality of care: Patient satisfaction with hospital noise levels at night, hospital staff helpfulness, and information received about home recovery significantly decreases by 5-10%, resulting in potential financial damages for hospitals. Third, disciplinary actions against doctors, physician assistants, and nurses increase by 34%. Supplementary analyses investigate admissions, cannabis-associated procedures (lung ventilation), hospital's colocation with recreational cannabis stores, and familiarity with cannabis patients.

4 - Telehealth in Acute Care: Pay Parity and Patient Access

Ozden Cakici, American University, Washington, DC, United States, Alex Mills

Several US states have recently adopted telehealth pay-parity policies. Such policies require that payers reimburse healthcare providers for telehealth services at the same rate as office visits. But health policy researchers have stated that telehealth may not be as effective as an office visit for acute care. Specifically, telehealth is associated with increased probability of a subsequent office visit (a "duplicate visit"). We examine whether
telehealth pay-parity policies are effective at improving access to acute care, and under what conditions. Using a three-stage game theoretic model, we study the impact of telehealth pay parity. Patients arrive and choose between telehealth and office visits according to an equilibrium queueing network. We find structural results for the equilibria under both homogeneous and heterogeneous illness types, and we fully characterize the equilibria in closed-form for homogeneous patients. When the duplicate visit rate is moderate, pay parity leads providers to allocate too much capacity to telehealth, resulting in lower overall patient access than could be otherwise achieved. We characterize a reimbursement level that avoids this misalignment and maximizes patient access, which is less than parity. On the other hand, copayments cannot align the system. In the fee-for-service environment that is common in the US for acute care, duplicate visits resulting from telehealth lead to an incentive alignment problem because they generate extra work and provider revenue, without any corresponding increase in patient access. Legislating pay parity for telehealth can lead to providers committing more capacity to telehealth, which may not always be good.

MC58

Summit - Terrace Suite 2

Responsible Optimization, ML, and AI in Healthcare

Panel Session

Health Applications Society

Co-Chair: Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Co-Chair: Gian-Gabriel Garcia, Georgia Institute of Technology, Atlanta, GA, United States

1 - Moderator Panelist

Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

2 - Panelist

Margret Bjarnadottir, University of Maryland, College Park, MD, United States

3 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

4 - Panelist

Vishal Ahuja, Southern Methodist University, Dallas, TX, United States

5 - Panelist

Fernanda Bravo, UCLA Anderson School of Management, Los Angeles, CA, United States

MC59

Summit - Ballroom 1

APS Distinguished Speaker: Dylan Foster

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Bridging Learning and Decision Making

Dylan Foster, Microsoft Research, New York, NY, United States

This tutorial will give an overview of the theoretical foundations of interactive decision making (high-dimensional/contextual bandits, reinforcement learning, and beyond), a promising paradigm for developing AI systems capable of intelligently exploring unknown environments and making decisions in a data-driven fashion. The tutorial will focus on connections and parallels between supervised learning/estimation and decision making, and will build on recent research which provides (i) sample complexity measures for interactive decision making that are necessary and sufficient for sample-efficient learning, and (ii) unified algorithm design principles that achieve optimal sample complexity. Using this unified approach as a foundation, the main aim of the tutorial will be to give a bird's-eye view of the statistical landscape of reinforcement learning and decision (e.g., understanding what modeling assumptions lead to sample-efficient algorithms). Topics covered will range from basic challenges and solutions (exploration in tabular reinforcement learning, contextual bandits) to the current frontier of understanding. We will also highlight practical algorithms.

MC60

Summit - Ballroom 2

Under the Roof Process Design and Optimization

Invited Session

The Practice Section of INFORMS

Chair: Herm Li, Amazon, Sammamish, United States

Co-Chair: Herm Li, Amazon, Sammamish, United States

1 - Amazon Logistics Labor Planning Platform and Its Eco System

Herm Li, Amazon, Bellevue, WA, United States, John Jiang

Amazon Logistics (AMZL) manages a network of delivery stations (DS) that are essential for its last-mile delivery services. Each station functions as a crucial hub, receiving packages from fulfillment centers, sorting them, and dispatching them to their final destinations. The efficiency of these operations relies significantly on the effective planning and utilization of associates working at the delivery stations. This presentation covers the design and implementation of a labor planning platform that optimizes various labor factors, including hiring, release, labor share, voluntary time off, and voluntary extra time, ensuring demand is met while minimizing labor costs. Over the years, modular extensions of the basic labor planning model have been developed, creating an ecosystem that addresses multiple aspects of labor planning. This includes the Mandatory Extra Time (MET) Estimator, Hourly Internal Transfer System (HITS), and Associate Schedule Generator (ASG). This presentation will also introduce these modules and demonstrate how the labor planning eco system coordinates AMZL's labor planning and influences operations.

2 - Experimental Design and Analysis in Supply Chain Operations

Brayan Ortiz, Amazon.com, Bothell, WA, United States, Andrew Bruce

Technology innovations are driving changes to fulfillment networks and supply chain operations. Sensors combined with sophisticated computer vision and machine learning algorithms offer the promise of increased automation and quality control. The technology, however, comes with significant cost and real-world performance often doesn't match offline controlled studies. Before making a significant investment, companies will conduct pilot studies to assess the benefit of the innovation in actual operations.

In contrast to traditional industrial experimentation (see "Statistics for Experimenters", Box, Hunter, and Hunter, 1978), field studies in supply chains share more in common with clinical trials in medical research. Pilot studies at Amazon typically involve multiple phases, with the initial phases focusing on safety and reliability and the subsequent studies designed to determine efficacy. Since pilots are conducted in operational settings, confounding factors can impact the outcomes. In particular, a main source of variation is often due to differences between individuals performance. It is important to design the study with sufficient statistical power to ensure reasonable confidence in the outcomes.

Amazon is developing a framework to conduct pilot studies addressing the challenges with experimentation in the field. This framework provides technical subject matter experts and product managers a guideline to design a statistically robust pilot study, from definition of the hypotheses to final analysis and outcome determination. Drawing from actual pilot studies at Amazon, we review key concepts in this framework, such as randomization and sample size determination, along with common pitfalls that experimenters often face in the field.

3 - A General Framework for Scheduling Under-the-Roof Processes in Amazon

Nikos Lappas, Amazon, Seattle, WA, United States

At Amazon, we harness the latest technological advancements and experiment with newly discovered capabilities, often resulting in diverse facility configurations, equipment types, and designs. Nevertheless, each facility must operate at maximum potential, which requires optimal resource scheduling to meet customer demand, on-time.

In this talk we will discuss how we leveraged the generality of the Resource-Task-Network (RTN) representation to adapt seamlessly to various facility floorplans & configurations, the importance of time representation and how the existence of a State definition unlocks multiple avenues for optimizing the schedule. We will also demonstrate how we applied concepts from Model Predictive Control (MPC) to develop a near-real-time adaptive scheduling approach, empowering facility managers to navigate the challenges and uncertainties of daily operations.

4 - From Science to Product: a Case Study on Labor Planning Platform of Amazon Logistics

Zhengqian jiang, Amazon, Santa Clara, CA, United States, Herm Li

Transforming science into a practical product is a complex journey. This talk will focus on how we implemented scientific models into products now utilized daily by Amazon's most critical logistics networks, aiding users in making crucial labor planning decisions. We will discuss three key aspects: selection of the optimal solution approach, ensuring the accuracy and reliability of data through cleansing and validation, and building user-friendly interface with effective data visualization. Through these aspects, we aim to address several challenges, including: the rapidly changing nature of labor planning operations, earning user trust, and demystifying the optimization model for users.

MC61

Summit - Ballroom 3

How to Start an Edelman-Worthy Project

Panel Session

The Practice Section of INFORMS

Co-Chair: Irv Lustig, Princeton Consultants, Princeton, NJ, United States

Co-Chair: Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, United States

1 - Moderator Panelist

Natalia Summerville, Memorial Sloan Kettering Cancer Center, Cary, NC, United States

2 - Panelist

Masoud Zarepisheh, Memorial Sloan Kettering, New York, NY, United States

3 - Panelist

Tugce Martagan, Eindhoven University of Technology, Eindhoven, Netherlands

4 - Panelist

Theodore Allen, Ohio State University, Columbus, OH, United States

5 - Moderator Panelist

Irv Lustig, Princeton Consultants, Princeton, NJ, United States

6 - Panelist

Jose Antonio Carbajal Orozco, iHeartMedia, San Antonio, TX, United States

7 - Panelist

Lei Wang, Optimization Analytics Technology, Singapore, Singapore

MC62

Summit - Signature Room

Interventions for Patients with Complex Medical and Social Needs

Invited Session

TutORial

Chair: Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

1 - Interventions for Patients with Complex Medical and Social Needs

Hari Balasubramanian, University of Massachusetts, Amherst, Amherst, MA, United States, Sindhoora Prakash, Ali Jafari, Arjun Mohan, Chaitra Gopalappa

Patients with multiple chronic conditions and social needs represent a small percentage of the population but have a disproportionate impact on healthcare costs and utilization. Organizations around the United States have created programs – often referred to as complex care interventions – to improve the health and well-being of such patients and reduce avoidable hospital and emergency department use. In this tutorial, we focus on two emerging themes in the field: (1) identifying clinically meaningful subgroups in complex care populations through unsupervised learning methods; and (2) describing the key operational features of interventions with an emphasis on staffing needs and the impact on patient outcomes. The material presented in this tutorial draws on the research of the Healthcare Operations Research Lab at the University of Massachusetts, Amherst, and its collaborating partners. To illustrate these themes and contextualize the details of complex care delivery, we use a range of patient-level examples, visualizations, descriptive summaries, case studies, and results from the clinical literature.

MC63

Regency - 601

Social Media Insights and Innovations: Unveiling New Trends and Perspectives

Invited Session

Information Systems

Chair: Hongyi Zhu, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Assessing Public Sentiment on AI-Generated Art Across the United States Using Deep Learning

Lan Sang, University of Colorado Boulder, Boulder, CO, United States, David Dobolyi, Kai Larsen

Art is crucial in conveying the rich tapestry of human feelings and societal values. However, the recent development of high-quality AI-generated art has created controversy in terms of how this new form of art should be perceived and valued. Understanding the public's attitudes towards AI-generated art can help AI companies improve products and guide the government's policy regarding the generative AI industry. This study uses social media data (i.e., Twitter/X) to analyze sentiment toward AI-generated artwork across the United States. Analyses of these data show a generally negative attitude towards AI art in most states. The findings also indicate that AI and art market maturity and urbanization rate positively impact public sentiment on AI-generated art, whereas median household income is negatively correlated with attitudes toward AI-generated artwork. Collectively, these findings highlight the importance of building mature AI and art markets.

2 - SKNews: A Design Framework for Skill-oriented Education Program Rankings

Yueqing Li, University of Arizona, Tucson, AZ, United States, Yong Ge, Shuojia Guo

Education program rankings are pivotal in shaping the reputation of higher education institutions and guiding academic choices for students and parents. While the labor market is currently experiencing significant transformations, students try to seek programs that enhance their competitiveness in the job market, expecting training that not only improves their skill sets but also aligns with employer demands. However, traditional rankings often fall short in providing the detailed insights necessary to meet these specialized needs. To bridge this gap, we introduce a deep learning framework that delivers personalized, reasonable, and demand-driven rankings of education programs. Our framework comprises three core modules: (1) Skill Extractor, which employs the Wide and Deep model to distill skill-related keywords from an extensive corpus; (2) CookieLM, a hybrid model that enhances a pre-trained language model with factual knowledge and domain-specific expertise to generate skill representations; and (3) ERIQ, a query-based ranking method that provides customized educational program rankings according to individual user requirements. The empirical analyses demonstrate our framework's superior effectiveness over existing techniques and offer insightful, user-specific evaluations across different scenarios.

3 - Evaluating Social Media Surveillance Risks: A Multimodal Large Language Model Approach to Linking Sensitive and Identifiable User Accounts

Fangyu Lin, The University of Texas at San Antonio, San Antonio, TX, United States

Social media surveillance poses a significant societal issue, with many users unaware of the risks associated with linking social media accounts that share sensitive information to identifiable ones. In this study, aiming to evaluate the likelihood of connecting accounts across social media platforms and increase user awareness of potential surveillance risks, we design a novel Cross-platform Social Media Surveillance Risk Assessment Framework. Central to this framework is X-FuseER, a multimodal Large Language Model (LLM)-based User Identity Linkage (UIL) model with two novel features: (1) Intermodal Attention that identifies and emphasizes modality-specific features critical for cross-modality relevance and (2) Hybrid Cross-modal Contrastive Loss that optimizes the similarity of embeddings within and across modalities for matched and unmatched record pairs. We systematically evaluated X-FuseER against benchmark ER and multimodal models on LinkSocial and self-developed UIL datasets. X-FuseER consistently outperformed benchmark models by statistically significant margins. To demonstrate the proposed framework's practical utility, we executed a case study to evaluate an anonymized Reddit user's social media surveillance risk. The findings reveal that the user's anonymized Reddit account, where she shares her explicit photos, can be linked to an identifiable Instagram profile. The proposed framework has important implications for various stakeholders, such as protest participants. In addition, we discuss several design principles of X-FuseER that can help guide future design science-based fraud detection, social media analysis, forensics and criminal investigations research.

4 - A Comparative Analysis of Pure and Shoppable Ads on Social Media Platforms

Jiayu Fan, Temple University, Philadelphia, PA, United States

Social media platforms are increasingly adopting monetization strategies that integrate e-commerce functionalities. Platforms like TikTok and Instagram have pioneered the use of shoppable ads, which can generate both ad revenue and commission revenue. For social media platforms, it is unclear whether shoppable ads are more profitable than pure ads. Furthermore, the ad pricing strategy and ad amounts can be different for pure ads and shoppable ads, which may be affected by the commission rate. However, there is no existing study that can answer these questions. Hence, our study fills this gap by providing a framework to compare the profit, ad price, and ad amount for pure ads and shoppable ads. Our findings suggest that shoppable ads may generate higher profit than pure ads even if users dislike shoppable ads more than pure ads and the product value sold through shoppable ads is relatively low. In addition, if users dislike pure ads more than shoppable ads: 1) the platform may still price pure ads at a lower rate compared to shoppable ads when the market size of pure ads is larger than that of shoppable ads; 2) the platform may put more pure ads than shoppable ads, interestingly, this may happen when users' aversion to pure ads is greater than a given threshold meanwhile product value is relatively high. This understanding provides actionable insights for platforms looking to optimize their advertising revenue and commission revenue.

MC64

Regency - 602

Harnessing the Power of AI and Analytics for Digital Platforms

Invited Session

Social Media Analytics

Chair: Sridhar Nerur, University of Texas-Arlington, Arlington , TX, United States

Co-Chair: Sagar Mahesh Badve, The University Of Texas At Arlington, Bedford, TX

1 - Is More Always Better? A Deep Dive into Twitch Streamer Success

Roopa Ramesh Desai, The University of Texas at Arlington, Arlington, TX, United States, Kay-Yut Chen, Jennifer Zhang

Video games have had a profound impact on entertainment and leisure activities since their inception in the 1950s. Technological advancements and the availability of various devices and platforms have increased video games' popularity across all age groups. Social media platforms like Twitch.tv, YouTube, and Facebook have led to the rise of video game live streaming as a profession, especially among young people. With the growth of the video game industry, live streaming has become a profitable profession for many streamers who focus on maximizing their revenue through various means. A video game streamer's effort level significantly impacts their revenue, partnership/sponsorship opportunities and audience growth. Thus, this paper investigates the relationship between streaming effort, streamer quality, and additional factors in the context of Twitch streaming, employing game theory as a theoretical background to further support the findings. We begin by examining how streamer effort, measured by hours streamed, is influenced by perceived streamer quality. A market share model is developed to explore the influence of streamer quality fluctuations on streaming hours and subsequent market reach. Using total subscribers as an initial proxy for quality, we uncover an unexpected concave relationship, suggesting that beyond a certain threshold, increased quality may lead to diminishing returns in the streaming effort. To further explore the components of streamer quality, we introduce additional variables such as age, involvement in esports, and gender. Our analysis explores various aspects that could potentially define streamer quality, shedding light on the complex dynamics of streamer success on Twitch.

2 - Illuminating the Shadows: Leveraging OSINT and AI for Enhanced Analysis of Dark Web Activities

sreehas Gopinathan, University of Texas Arlington, Arlington, TX, United States, Sridhar Nerur

The dark web, a secluded fraction of the internet often associated with anonymity and illicit activities, presents significant challenges for cybersecurity professionals and law enforcement agencies. The traditional approaches to monitoring and analyzing these hidden services are often hampered by the inherent complexities of the dark web's architecture and the vast amount of unstructured data it contains. This presentation proposes a novel framework that harnesses the power of Open Source Intelligence (OSINT) and Artificial Intelligence (AI) to enhance the visibility and understanding of dark web ecosystems.

Our approach utilizes advanced AI techniques, including machine learning algorithms and natural language processing, to systematically collect and analyze data from the dark web. By integrating OSINT tools, our method not only identifies and aggregates data from publicly accessible dark web sources but also intelligently analyzes content to uncover patterns, trends, and actionable intelligence. This framework is designed to provide a comprehensive analysis of dark web marketplaces, forums, and other platforms, facilitating the detection of illegal activities, such as drug trafficking, weapon sales, and cyber threats.

3 - AI-Assisted Distributed Cognition in Data Science

Kaiyue Liu, University of Texas at Arlington, Arlington, TX, United States, Jayarajan Samuel, Sridhar Nerur

With the proliferation of generative AI, the debate about whether AI functions merely as a tool or as a collaborative partner continues. Although AI's capabilities recently improved significantly to solve increasingly complex problems, empirical examination of how AI collaboration impact task performance remains limited. Our study applies Distributed Cognition Theory (DCT) to examine the interaction between an artificial cognition system (ChatGPT) and an external structural cognition system (mind maps) on task performance, cognitive load, job satisfaction, and self-efficacy. Utilizing an experimental design, this research aims to peek into the mechanisms through which AI affects task performance. We anticipate demonstrating that AI usage improves performance and decreases cognitive load, with these effects significantly moderated by external cognition systems. This study aims to fill a literature gap in the empirical examination of AI's impact on productivity and cognitive dynamics. The findings are expected to challenge AI's role as merely auxiliary tools, proposing that effective AI integration enhances both individual and collaborative cognitive functions. This may provide valuable insights for practitioners on optimizing AI use and pave the way for future research into the cognitive dynamics of human-AI collaboration, potentially reshaping how AI is implemented in professional settings.

4 - Impact of Generative AI on the Value of Peer Produced Content - Evidence from Wikipedia

Srikar Velichety, University of Memphis, Memphis, TN, United States

Generative AI (GenAI) technologies such as ChatGPT, Gemini, Alpha Code, CoPilot, Diffusion, Dall-E, etc., have made headlines since late 2022 due to their ability to easily generate content, from text to multimedia, in various domains. These technologies are improving the productivity of workers in multiple industries, especially those that are based in the knowledge economy. Moreover, they are consistently reducing factual-content errors and making information easily accessible through multiple information formats including audio, images, and text. Large language models are at the core of this GenAI revolution. They are trained using publicly available information from platforms such as Wikipedia, Stack Overflow, GitHub, etc. Although GenAI tools have made information search more efficient, recent research shows they are undermining and degrading engagement with online question and answer (Q&A)-based knowledge communities like Stack Overflow and Reddit. We extend this stream of research by examining the impact of GenAI on the market value and quality of peer-produced content using the world's largest open-source knowledge repository i.e., Wikipedia, which is different from Q&A-based communities mentioned above. We follow an approach like Burtch et al. (2023) and extend empirical analyses focusing on ChatGPT's release on November 30, 2022. We collect monthly Wikipedia page views and content (text) data for six months before and after the release date as the treatment group. We then collect data for same months a year before as the control group. The difference-in-difference (DID) analyses demonstrate significant decrease in Wikipedia page views (market value) after the release of ChatGPT.

MC65

Regency - 603

Teaching Data Analytics with Low Code Tools

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Angelika Leskovskaya, SMU Cox School of Business, Dallas, TX, 75275, United States

1 - Panelist

Graeme Warren, Johns Hopkins University, Washington, DC, United States

Teaching MS Business Analytics and Risk Management using Microsoft's Responsible AI (RAI), h2o's autoML and/or Explainable AI (XAI) toolkit, PyCaret (autoML)and AnyLogic (discrete event models only) in his operations management course (for Flex MBA students).

2 - Panelist

Angelika Leskovskaya, SMU Cox School of Business, Dallas, TX, United States

Teaching Predictive Modeling and ML using ForecastX, Alteryx for undergraduate BASCM program at SMU Cox School of Business.

3 - Panelist

Deanna Sanchez, Nova Geographica LLC, Denton, TX, United States

Alteryx ACE and Owner, Nova Geographica, LLC, Senior Instructor and Developer of Alteryx Beginner, Intermediate and Advanced Training; Advanced spatial and demographic analytics, utilizing Alteryx and various GIS (Geographical Information System) softwares such as ESRI ArcGIS, MapInfo, and Qlik GeoAnalytics. Enterprise GIS / IT implementation, automation and management. Alteryx Analytic Gallery administration and custom App creation; Alteryx reporting, map output and process automation.

4 - Panelist

Kyle Maclean, Ivey Business School, London, ON, Canada

Teaching Sports/Entertainment Analytics using Jamovi (an open-source and free platform built on R but is no-code).

5 - Panelist

Michael Veatch, Gordon College, Wenham, MA, United States

What I teach: introductory statistics, mathematical statistics, optimization (LP, IP, network opt, NLP), some data wrangling Audience: undergraduate math, data science, computer science, economics, and science majors Low code tools: R (taught with templates that greatly reduce coding), Excel Solver, AMPL, CPLEX, and a little optimization software that goes with my book.

MC66

Regency - 604

Emerging Problems in Platform Operations

Invited Session

Artificial Intelligence

Chair: Ruohan Zhan, The Hong Kong University of Science and Technology, Clear Water Bay, N/A, Hong Kong

Co-Chair: Zhengyuan Zhou, Stern School of Business, New York University, New York, NY, United States

1 - Assessing External Validity of A/B Tests Under Effect Ordering

Ariel Boyarsky, Columbia University, New York, NY, United States, Hongseok Namkoong, Naoki Egami

Assessing the external validity of experimental results is a pervasive challenge in the experimental design literature, threatening the generalizability of internally-valid findings. External validity issues, well-documented in the A/B testing and distribution shift literature, are prevalent in clinical trials, online platforms, and social science experiments. While specifying a target population is a common strategy to address these concerns, practitioners often lack access to the complete target distribution. To bridge this gap, we propose an optimization-based framework enabling practitioners to define target distributions at the site-level. To address common collinearity issues inherent in site-level experiments, we utilize the effect ordering principle, in which direct effects outweigh interaction terms, to obtain partially identified sets for the treatment effect on the target population. Our methodology employs a hierarchical semiparametric model, where potential outcomes are influenced by direct and indirect nuisance parameters estimated from individual-level covariates using machine learning techniques. We introduce an augmented debiased estimator and cross-fitting algorithm, and demonstrate that it achieves central limit behavior. Moreover, our optimization procedure facilitates flexible implementation of variations of the effect-ordering principle, tailored to specific domains. To illustrate the effectiveness of our approach, we apply it to the micro-credit expansion experimental dataset from Meager (2019) and survey experiment data from \Naumann (2018).

2 - Llm and Experimentation

Zikun Ye, University of Washington, Seattle, WA, United States, Hema Yoganarasimhan, Yufeng Zheng

We integrate LLM-based prediction models with bandit algorithms to enhance news headline selection. Our methodology encompasses several strategies, including in-context learning, GPT embedding-based classification, and LLM fine-tuning. These approaches achieve over 80% accuracy in headline selection. By pairing the prediction model with the 2UCB algorithm, we facilitate effective online decision-makingsignificantly boosts revenue.

3 - Adaptive Instrument Design for Indirect Experiments

Vasilis Syrgkanis, Stanford University, Stanford, CA, United States, Yash Chandak, Shiv Shankar, Emma Brunskill

Indirect experiments provide a valuable framework for estimating treatment effects in situations where conducting randomized control trials (RCTs) is impractical or unethical. Unlike RCTs, indirect experiments estimate treatment effects by leveraging (conditional) instrumental variables, enabling estimation through encouragement and recommendation rather than strict treatment assignment. However, the sample efficiency of such estimators depends not only on the inherent variability in outcomes but also on the varying compliance levels of users with the instrumental variables and the choice of estimator being used, especially when dealing with numerous instrumental variables. While adaptive experiment design has a rich literature for direct experiments, in this paper we take the initial steps towards enhancing sample efficiency for indirect experiments by adaptively designing a data collection policy over instrumental variables. Our main contribution is a practical computational procedure that utilizes influence functions to search for an optimal data collection policy, minimizing the mean-squared error of the desired (non-linear) estimator. Through experiments conducted in various domains inspired by real-world applications, we showcase how our method can significantly improve the sample efficiency of indirect experiments.

4 - A Simple Formulation to Causal Clustering

Jinglong Zhao, Boston University, Boston, MA, United States

We consider a network interference problem, in which an experimenter conducts experiments over a single connected network. We consider the setting where the network is known and the exposure mapping, which describes how one unit's treatment assignment impacts another's potential outcomes, is well-specified. In this setting, causal clustering refers to conducting clustered experiments to achieve accurate estimation and inference. This paper proposes a simple robust optimization approach to causal clustering, which also synthesizes several existing results in the literature.

MC67

Regency - 605

AI, Operations, and Platforms

Invited Session

Information Systems

Chair: Manmohan Aseri, Robert H. Smith School of Business at University of Maryland, Pittsburgh, PA, United States

1 - Platform Intermediation and Firm Investments

Vaarun Vijairaghavan, University of Calgary, Calgary, AB, Canada, Barrie Nault

We study how firm investments can be coordinated in a setting where firms offering two kinds of complementary technology operate on a platform that intermediates the relationship between firms. We model this coordination problem by abstracting from the case of Demand Response (DR) on the Smart Grid, where the platform enables interoperability between two distinct but complementary types of DR products, firms invest to develop DR products, customers purchase DR products from firms, and a policymaker favors greater customer adoption of DR. In our setting, demand generated by a firm depends on own investments and prices which they choose, as well as cross-side investments and prices by other firms. The platform sets fees and transfers between firms. Among other results, we find that the platform and the policymaker goals are aligned as far as the transfers are concerned: the platform's choice of transfer coordinates firm prices and investments, and maximizes aggregate platform demand.

2 - Shipping Consolidation Using AI for Different Delivery Deadlines and Modes

VARUN GUPTA, University of North Georgia, Dahlonega, GA, United States

Shipping costs are a major cost to a company and many logistics service providers (e.g.; 3PL providers) promise significant savings on shipping costs to their customers. We explore how 3PL providers can utilize historical shipping data from their customers to implement shipping consolidation in real-time using advanced AI methodologies.

3 - Social Learning and Content Quality Under Polarization

Bharadwaj Kadiyala, University of Utah, Salt Lake City, UT, United States, Dongwook Shin

We study how polarization influences content consumption and production on digital platforms that monetize consumer engagement. Specifically, we consider a content that advocates a particular position on a divisive issue. Consumers with polarized preferences towards the content's position are sequentially exposed to the content. Initially, consumers are uncertain about the content quality, but they have the opportunity to learn about it using aggregate consumption metrics and other informative signals. We find that under polarization, social learning based on consumption metrics can mislead consumers to perceive low-quality content as higher quality, even in the long run. We explore the implications of this result on content design and platform operations.

4 - Ai-Generated Content: Recommend OR Not?

Di Yuan, University of Illinois Urbana-Champaign, Champaign, IL, United States, Luying Wang

With the increased usage of generative AI tools in digital content production, platforms adopt different approaches toward distributing AI-generated content. For example, Google demotes web pages written with generative AI in its search results, while Facebook and Instagram plan to assign a label to images and posts identified as AI-modified. Such divergent attitudes toward AI-assisted content reveal an intriguing question about optimal platform design with the rising trend in generative AI usage. We devise an analytical model to explore the economic implications of various platform policies on AI-assisted content detection and distribution.

Our model considers a monopoly platform with heterogeneous creators, where the creators' decision to use AI has implications for both horizontal and vertical competition. The platform, in turn, strategically decides whether to detect the creators' usage of AI and, subsequently, whether to recommend or hide it. By examining the strategic interactions between the platform and creators, we offer insights into the complex dynamics at play in the realm of AI-assisted content production and distribution.

MC68

Regency - 606

Revolutionizing Healthcare: Advanced AI and Data Analytics for Decision Support

Invited Session

Analytics Society

Chair: Bahareh Kargar, New Jersey Institute of Technolog, Newark, NJ, United States

1 - A Mixed Integer Linear Programming Approach for Nurse Scheduling Problem with Diverse Read-World Factors

Hansol Kim, Department of Industrial Management Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Gyeongbuk, Korea, Republic of, Eun-guk Cho, Byung-In Kim, Jinsuk Lee, Donghyung Kang

The nurse scheduling problem (NSP) holds significance in healthcare management. Currently, nurse rosters are predominantly manually created in hospitals, presenting significant challenges and consuming considerable time. In most Korean hospitals, head nurses manually draft rosters for each ward, undergoing multiple revisions before finalization, a process taking a minimum of 6 hours per ward. These rosters must satisfy the requests of nurses and adhere to scheduling rules. Automating this process offers the potential to generate higher-quality schedules within a more efficient timeframe. In this study, we propose a mixed integer linear programming approach designed to efficiently generate monthly schedules for 20 to 50 nurses, aligning with scheduling rules and meeting nurses' preferences. Our approach has been successfully deployed across 20 wards at Pohang SM Christianity Hospital, validated for its effectiveness, and is actively utilized in scheduling operations.

2 - A Deep Learning Approach to Predict Patient Length of Stay in a Burn Unit

Bryan Norman, Texas Tech University, Lubbock, TX, United States, Rafael Cacao

The length of stay (LOS) for patients in a burn ward is a crucial aspect impacting healthcare delivery, patient care quality, clinical management, and resource allocation. Despite its significance, LOS estimation methods have predominantly relied on a regression model dating back to 1987, with subsequent updates in studies from 2011. In this traditional regression model, Total Body Surface Area (TBSA) affected by burns has been regarded as the primary determinant of LOS. Patients with larger TBSA commonly experience prolonged LOS periods alongside heightened mortality rates. However, LOS can exhibit significant variations, even among patients with similar TBSA percentages. Additionally, burn patients frequently present with a spectrum of complex concurrent injuries, encompassing fractures, soft tissue trauma, traumatic brain injuries, infections, acute kidney injury, ventilator-associated pneumonia, and various comorbidities. These factors further complicate LOS prediction. In this study, we introduce a deep learning model that utilizes admission data to forecast a patient's LOS. Preliminary results demonstrate a notable improvement in prediction accuracy, with an R-squared value of 0.82, surpassing that of current state-of-the-art prediction models.

3 - Implementation of a Machine Learning Tool into the Electronic Health Record to Predict Delirium: An example illustrating MLOps in Medicine

Shant Ayanian, Mayo Clinic, Rochester, MN, United States, Sandeep Pagali

Background

Multicomponent delirium prevention tools have been successful in reducing episodes of delirium in hospitalized patients. Previously published models have proven effective in a prospective study. We describe the implementation of delirium prediction tool into clinical practice.

Purpose

Having a tool functioning seamlessly within the EHR and reporting back to providers and nurses in real time would be beneficial for patients as targeted interventions would reduce their risk for delirium.

Description

The delirium prediction tool is a set of two elastic net logistic regression models designed for medical and surgical patients, respectively.The implementation was challenging as two different models had to run separately on medical and surgical patients and be reported back. The two models were uploaded into an EHR native cloud service where they will receive live data feeds and provide a probability of delirium on a scale of 0-1 every 12 hours. Physicians see these probabilities on their patient lists and nurses within the nursing documentation flowsheets. The probabilities are also color coded to indicate low, intermediate, or high probability. Providers use that information to consult a geriatrics service for recommendations and managing medications. An ongoing pilot project focused on nursing workflows and its results will be presented in this poster as well.

Conclusions

This poster will highlight the feasibility of implementing an AI (Artificial Intelligence) model within the EHR and integrating it within existing workflows to benefit patients and providers.

4 - From Algorithms to eLearning: The Rollercoaster Ride of Digital Integration and Sustainment in Healthcare

Lindsey Philpot, Mayo Clinic, Rochester, MN, United States

Advanced analytics, predictive algorithms and other digital tools hold potential to transform healthcare delivery. Implementation and sustainment of these tools in clinical practice settings can be challenging. From an implementation and healthcare delivery science perspective, three case examples will be discussed considering successes and failures of integrating digital tools into practice: an algorithm for identifying patients with palliative care needs, a tool to help recommend imaging and procedural services for patients with advanced needs, and a suite of eLearning solutions for patients experiencing chronic pain.

5 - Equitable resource allocation with a health care application: is equity in the eye of the modeler?

Emanuele Blasioli, McMaster University, Hamilton, ON, Canada

This study examines equitable resource allocation by introducing four equity constraints categorized into general and domain-specific criteria. An equitable resource allocation strategy seeks to distribute a finite amount of resources with the goal of optimizing an objective function while complying with one or more equity principles. This can be accomplished by customizing the allocation process based on the criteria created to meet the unique requirements of the stakeholders. The need to ensure equity, in terms of serving populations or regions according to established equity standards, and to be efficient in terms of maximizing resource utilization for peak performance often poses a challenge. This challenge is analyzed both mathematically, as a trade-off, and in the context of policy-making. We represent equity through inequality-based indices, due to their compatibility with the knapsack model for addressing resource allocation issues. This work provides guidance on how equity interacts with effectiveness in an optimization framework. We illustrate our proposed methodology on a COVID-19 vaccine allocation case study and provide related managerial insights.

MC69

Regency - 607

Yu Kan's Session

Invited Session

eBusiness

Chair: Yu Kan, University of Washington- Michael G. Foster School of Business, Seattle, WA, United States

1 - Taste of Free Lunch: The Impact of Free Product Disclosure on Review Dynamics

Yimin Yin, University of Hong Kong, Hong Kong, Hong Kong, Yipu Deng, Xiao Lei

This study investigates the impact of free product disclosure on review platforms, where reviewers voluntarily reveal if they received the product for free. Specifically, we examine whether and how this disclosure policy influences subsequent review generation. Using a quasi-experimental design on Steam, we find that free label reviews increase both volume and efficiency of subsequent reviews. Notably, the increase in volume is more pronounced when free label reviews are of poorer quality, indicating a disconfirmation effect. Additionally, active reviewers, who tend to write effective reviews, contribute significantly to review efficiency under this effect. Our findings suggest that user-generated content (UGC) platforms could leverage free product disclosure as a cost-effective strategy to improve subsequent UGC. This study also complements prior research showing that disclosing the exchange of benefits may not rectify the bias from benefited reviewers; instead, given proper disclosure, the UGC community can organically address this bias through subsequent reviews.

2 - Fleeting or Lasting Success? Unlocking the Power of Livestream Selling

Yipu Deng, University of Hong Kong, Hong Kong, Hong Kong, Weilong Wang, Guoxin Li, Jinyang Zheng, Karthik Kannan

Livestream selling has emerged as a powerful marketing tool for product promotion, delivering substantial revenue growth for retailers. This study delves into the cost-effectiveness of livestream selling by examining its temporal impact on sales during post-adoption days without live streaming, and seeks to unravel the mechanisms underpinning its success. Collaborating with a large Asian streaming platform, we employ a quasi-natural experimental setting. Our findings reveal a significantly positive immediate impact on sales while no discernible lagged effects are observed. This short-lived impact indicates that real-time interactions play a crucial role in driving performance within frequent livestream selling contexts. By offering a comprehensive understanding of the temporal dynamics of livestream selling, our study not only enriches the existing literature but also provides valuable insights for industry practitioners.

3 - Optimizing Paywall Design in Freemium Along with Pay-Per-Use Premium Access

Junzhi Xue, University of Science and Technology of China, Hefei, China, People's Republic of, Lizheng Wang, Yongjun Li, Yong Tan

Paywall has emerged as a pivotal monetization strategy for digital content platforms. In this study, we investigate the paywall design in the context of freemium along with pay-per-use access to premium, in which, consumers can have access to some free samples of a product and then pay for the remaining premium content per use. Based on the case of online book reading, we build a structural econometric model of how readers decide to keep purchasing the next chapter or stop reading forever by depicting consumer's book quality perception and forward-looking behaviors. We utilize detailed consumer chapter-by-chapter payment data to estimate the model. Through policy simulations, we reveal the optimal strategy of deciding the number of free samples and chapter price and verify the effectiveness of the nonuniform pricing strategy in the pay-per-use stage. We also find the optimal strategy for balancing consumer engagement and platform revenue. This study contributes to the literature on paywall design, pay-per-use pricing and engagement-revenue trade-off, offering valuable managerial insights for pricing content products in digital content industries.

4 - Customer Satisfaction as a Moderator: Unpacking the Effects of Managerial Response on Future Review Engagement

Yuqing Hu, University of Washington, Seattle, WA, United States, Lin Jia, Tongxin Zhou, Yong Tan

Online managerial responses not only influence the customers who receive them, but also are kept online as a public reference, affecting future engagement of potential customers. Existing literature presents inconsistencies regarding the impact of responses and the optimal response strategy, leaving it unclear whether responding to positive, negative, or both types of reviews is more effective. We address this gap by proposing customer satisfaction as an important moderator, where we focus on the satisfaction of potential customers. Drawing on data from a leading travel platform in China, our findings confirm that customer satisfaction moderates the effectiveness of responses on future engagement. Responses to positive reviews are more effective when customer satisfaction is low, while responses to negative reviews are more effective when customer satisfaction is high. Our findings highlight that an effective response strategy should be contingent on the business's capability and underscore customer satisfaction as a critical strategic variable.

5 - The Effects of Subway Transit Expansion on Demand for Ridesharing Services

Juan Qin, USTC, Hefei, China, People's Republic of, Stephanie Lee, Yong Tan, Xusen Cheng

As city planners and policymakers formulate transportation policies and make urban planning decisions, it is imperative to understand the relationship between the subway transit system and ridesharing services, especially in current times when the demand for ridesharing services surged dramatically. In particular, when making social overhead capital (SOC) investments in urban transportation, such as investments in the expansion of subway lines, it is important to consider the nuanced interplay between ridesharing and public transit systems. In this paper, we examine how ridesharing services complement and work together with the subway transit system when new subway lines open. We exploit a natural experiment setting in which additional subway lines opened in Beijing, China, and utilize detailed ridesharing order information from a leading Transportation Network Company (TNC). By exploiting a natural experiment setting and employing the difference-in-differences framework, we find that ridesharing services complement the subway transit system. After the opening of new subway stations, the demand for ridesharing services traveling to and from subway stations significantly increases. Moreover, ridesharing services not only provide first-mile and last-mile connections to public transits for residents living close to subway stations, but also provide longer distance connections for residents living far away from subway stations. In addition, we find that the complementary relationship dominates for incumbent subway stations whereas the substitution relationship dominates for newly opened stations. We find that newly opened subway stations generate ripple (spillover) effects on existent subway stations by changing the underlying network topological features of the subway system.

MC70

Regency - 701

Next-generation Energy Markets

Invited Session

ENRE: Electricity

Chair: Daniel Bienstock, Columbia University, New York, NY, United States

1 - Next-Generation of Electricity Markets

RICHARD O'NEILL, self-employed, Silver Spring, MD, United States

We look look at the expected evolution of electricity markets and offer suggestions of efficiency enhancements including pricing, demand bidding, transmission planning, and the principle of beneficiaries pay/cost allocation.

2 - Resource Adequacy Under Deep Decarbonization

Conleigh Byers, Harvard University, Cambridge, MA, United States

The rise of variable renewable energy and storage in energy systems has called into question whether current electricity market designs are sufficient for a decarbonized future. While energy-only markets with locational marginal prices can theoretically provide cost recovery for the socially optimal mix of resources in the long-run, many systems today have turned to some form of centrally administered resource adequacy mechanism. In a future with high shares of zero-marginal cost resources, the positive skewness of revenue is expected to increase, i.e., capital costs will be recovered from fewer scarcity hours. This poses challenges for risk management, especially with incomplete markets in risk, exacerbating concerns about the ability of short-run markets to achieve resource adequacy.

Discussion has turned to increased reliance on centralized procurement, e.g., via hybrid markets. An alternative market-based approach to managing risk is to require a subset of market participants to enter into mandatory forward energy contracts. An advantage of this approach is that it could provide sufficient hedging such that regulators are more comfortable with allowing the volatility of full-strength energy prices that provide optimal operational and investment signals. The challenge of the contract design is providing hedging while preserving marginal incentives.

We formulate a mixed complementarity problem for incomplete markets and solve via a non-algorithmic approach that allows for exploration of multiple equilibria. We use this model to explore worst-case equilibria for reliability, allowing for a comparison of the consumer welfare from a resource adequacy perspective under different contacting regimes and levels of demand elasticity.

3 - Revealing Decision Conservativeness Through Inverse Distributionally Robust Optimization

Zhirui Liang, Johns Hopkins Univeristy, Baltimore, MD, United States, Qi Li, Andrey Bernstein, Yury Dvorkin

This talk introduces Inverse Distributionally Robust Optimization (I-DRO) as a method to infer the conservativeness level of a decision-maker, represented by the size of a Wasserstein metric-based ambiguity set, from the optimal decisions made using Forward Distributionally Robust Optimization (F-DRO). By leveraging the Karush-Kuhn-Tucker (KKT) conditions of the convex F-DRO model, we formulate I-DRO as a bi-linear program. We demonstrate that I-DRO not only guarantees the existence and uniqueness of an optimal solution but also establishes the necessary and sufficient conditions for this optimal solution to accurately match the actual conservativeness level in F-DRO. Furthermore, we identify three extreme scenarios that may impact I-DRO effectiveness. Our case study applies F-DRO for power system scheduling under uncertainty and employs I-DRO to recover the conservativeness level of system operators. Numerical experiments based on an IEEE 5-bus system and a realistic NYISO 11-zone system demonstrate I-DRO performance in both normal and extreme scenarios.

4 - Équilibria with Virtual Bidders

Garud Iyengar, Columbia University, New York, NY, United States

In wholesale electricity markets, suppliers and load-serving entities (LSEs) enter into contracts to secure energy in the Day-Ahead (DA) market and later adjust for any discrepancy between contracted quantities and actual deliveries in the Real-Time (RT) market. This two settlement structure offers hedging benefits for both suppliers and LSEs but may lead to market inefficiencies. We develop a supply function equilibrium model to explore how virtual trading — a financial strategy that allows participants to speculate on price differences between the

DA and RT markets without physical transactions — can mitigate these inefficiencies. Our analysis reveals that without virtual trading, LSEs often bid below their actual demand in the DA market, leading to DA prices that are lower than the expected RT prices. We show that introducing virtual trading narrows this price disparity and, with an increasing number of virtual traders, potentially eliminates it. Nonetheless, this mechanism prompts LSEs to further reduce their bids in the DA market, deviating from their true demand forecasts. Our analysis indicates that the presence of renewable energy suppliers leads LSEs to submit higher bids in the DA market, compared to scenarios with traditional energy suppliers only. We provide empirical support to our main model implications using data from the California and New York Independent System Operators.

5 - Market Equilibria and Efficiency for Linked Green-Grey Power Pools

Siyuan Wang, Argonne National Laboratory, Argonne, IL, United States, Muireann Lynch, Ben Hobbs

Concerns over climate change have led to electricity markets with greatly increased levels of variable renewable energy integration. New electricity market pricing mechanisms are needed for future highly renewable penetrated grids, to address challenges from resources with high capital cost and zero short-run marginal cost. In this work, we model a new market pricing approach in the context of the green power pool (GPP), which takes the form of an exclusive market for renewable power, cleared alongside a conventional power market. Two-way trading between the two markets is allowed. The definition, calculation methods, and properties of capacity pricing for renewable participating in the GPP are systematically presented. Using equilibrium models, we show that under the proposed pricing the GPP achieves the same market efficiencies with conventional market design in deterministic and risk-neutral stochastic settings. When renewable investors are risk-averse, compared to the conventional market, the proposed pricing can lead to different equilibria under a GPP, with an increased incentive for renewable generation.

MC71

Regency - 702

Modern High-Dimensional Inference - Theory, Methods and Applications

Invited Session

Data Mining

Chair: Trambak Banerjee, University of Kansas, Lawrence, KS, United States

1 - A Bootstrap Test for Equality of Covariance Matrices in High-Dimension

Nilanjan Chakraborty, Missouri University of Science and Technology, Saint Louis, MO, United States, Sayar Karmakar, Hira Koul

In this talk, we propose a test for testing the equality of two high-dimensional covariance matrices in the two-sample set up. This test is based on the maximum of the absolute differences between the entries of the multiplier bootstrap Jackknifed estimators of the two population covariance matrices. We proved a Gaussian approximation result for two-sample U statistics and used it to derive the asymptotic distributions of the sequence of the proposed test statistics under the null and some local alternative hypotheses. These results are obtained under some weak conditions on the moments of the random vectors and the tails of the marginal distributions. The correlation structures of the random vectors can be arbitrary, the two sample sizes need not be equal and the dimension is allowed to grow exponentially with the two sample sizes. The test is shown to be consistent against a class of shrinking nonparametric alternatives. A finite sample simulation study reveals some superiority of this test compared to some of the existing tests.

2 - Rank Based Tests for Heterogeneous Data

Ritwik Sadhu, Cornell university, Ithaca, NY, United States, Nilanjan Chakraborty, Trambak Banerjee

In many modern inference tasks, the data-generation process contains inherent sources distribution heterogeneity, often resulting in multi-modality of the composite dataset on which statistical procedures must be performed. Such data containing distributional heterogeneity allows testing for the presence of an entirely new component distributions in a subset of the data. For this testing problem, henceforth called the remodeling problem, we propose a new test statistic constructed using optimal transport (OT)-based multivariate ranks, which allows for asymptotically consistent testing with correct asymptotic level under the assumption of well separated component distributions in the null. To the best of our knowledge, this is the first statistic in the literature for this problem with the aforesaid guarantees.

3 - Statistical Inference for Subgraph Densities Under Induced Random Sampling from Network Data

Ayoushman Bhattacharya, Washington University in St. Louis, St. Louis, MO, United States, Nilanjan Chakraborty, Soumen Lahiri

In this talk, we develop a framework for obtaining statistical guarantees for subgraph densities of a general population network under without replacement sampling (SRSWOR). The examples of such subgraph densities include edge density, triangle density, two-star density and other popularly studied graph summary statistics. Under this sampling scheme, we obtained a Berry-Esseen bound to establish asymptotic normality of the Horwitz-Thompson (HT) estimator for the population subgraph densities. To facilitate inferential procedures, we provide a jackknife estimator of the unknown population variance and establish its consistency. Our method works under certain popular super population network models like sparse graphons, stochastic block models (SBM) and others. We also establish the joint asymptotic normality of two subgraph densities which is crucial in establishing asymptotic normality of the global clustering coefficient/global transitivity of the sampled graph. Our results find a useful application to the problem of testing the equality of two population graphs using the subgraph densities as the test statistic. Finally, we present a simulation study and a real data analysis which corroborates our theoretical findings.

4 - On a Fast and Consistent Test for Equality of Means for High-Dimensional Data

Sayan Das, Washington University in St. Louis, St. Louis, MO, United States, Debraj Das, Subhajit Dutta

In this talk, we propose a two-sample test for high-dimensional means using a logistic regression framework. The crux of our method relies on identifying significant feature variables via LASSO (using the logit link) that can capture the population mean. The proposed test is based on the regression coefficients of these selected features, effectively reducing dimensionality. Theoretically, we have shown that our proposed test is consistent against a wide range of alternatives. In the ultra-high-dimensional regime, our test leverages the sparse structures of the means, making it computationally more efficient compared to other methods in the existing literature. Additionally, we have extended our method to test for equality of means across multiple groups. Finite sample studies corroborate our theoretical findings and reveal the superiority of our tests compared to several existing tests.

MC72

Regency - 703

Expanding AI Boundaries to Education, eCommerce, and Sustainable Urban Mobility

Invited Session

Data Mining

Chair: Dervis Ozay, The University of Texas at Arlington, Arlington, TX, 76155, United States

Co-Chair: Shouyi Wang, The University of Texas at Arlington, Arlington, TX, 76019, United States

1 - A Novel Deep Learning Approach for Dynamic Pricing Optimization and Smarter Business Decisions: a Case Study of eCommerce

Dervis OZay, The University of Texas at Arlington, Arlington , TX, United States, Shouyi Wang, Bharani Nammi

The complexity of pricing strategies has become increasingly significant in the rapidly evolving world of e-commerce for companies looking to achieve profitability and remain competitive. There is perpetual flux in the pricing landscape, marked by swift competitor reactions, variable market conditions, and the requirement to set prices for a wide variety of products. Given this BSS Data Challenge opportunity, we developed a novel multi-step DL framework to provide optimal solutions to the dynamic pricing problem. In particular, we consider both Pricing and Ads Spend are decision variables. We propose two deep learning (DL) models. The first DL model is developed to make probabilistic forecasting and generate expected value ranges for Unit Price, Ads Spend, and Competitors' Prices given historical market behaviors. The second DL model is developed to model the complex relationships from Unit Price, Ads Spend, and Competitors' Prices to Unit Ordered and resulting Profit. The Deep Learning Model 2 is then used to predict the expected profit given different values of Unit Price and Ads Spend within the predicted ranges by the Deep Learning Model 1. The optimal Unit Price and Ads Spend are determined by their values that maximize Deep Learning Model 2's expected profit. The proposed DL framework is flexible and can be used to set daily optimal price and ads spend, as well as other time resolutions, such as every N days or one week. The proposed DL approach provides a novel data-driven solution to guide price and ads spend decisions for e-commerce products.

2 - Taming the AI Dragon: Reimagining Governance Frameworks and Data Management for AI Alignment

Darshan Desai, Pedago, New Providence, NJ, United States, Ashish Desai

As Artificial intelligence (AI) systems advances, it is becoming crucial to develop innovative governance structures and approaches to align these systems with human values early on. This research explores AI alignment challenges, focusing on large language models (LLMs), which present unique difficulties due to their generative capabilities and potential for bias. These models can produce unpredictable outputs, raising ethical concerns and the risk of perpetuating biases. Through strategic analysis of technology roadmaps, we investigate how governance frameworks can be reimagined to support AI alignment. We propose a novel framework for adaptive AI governance, emphasizing continuous oversight, high data quality, innovative architectural approaches, and stakeholder engagement. Our findings highlight the critical role of data management and governance in ensuring AI systems' reliability and trustworthiness, particularly in mitigating biases and ensuring more intentional augmentation and automation in decision-making processes. This study contributes to aligning AI technologies with ethical standards and societal values.

3 - Multiple Risks in Extended Markov Ratio Decision Processes

Alexander Zadorojniy, IBM Research, Haifa, Israel, Orit Davidovich, Takayuki Osogami

An Extended Markov Ratio Decision Process (EMRDP) enables the integration of a risk term into its underlying Markov Decision Process (MDP). We propose extending EMRDP to accommodate multiple risks and present an algorithm for determining the optimal policy for this enhanced EMRDP. We analyze both the theoretical properties of the algorithm and its empirical application to a grid world problem. Furthermore, we illustrate the application of EMRDP to financial trading, demonstrating that we achieve a strongly polynomial algorithm for this particular problem.

MC73

Regency - 704

Mining for Health: Data Mining for Personalized Care and Operational Efficiency

Invited Session

Data Mining

Chair: Sara Nourazari, California State University, Long Beach, Long Beach, CA, United States

1 - Comparing Retrieval Augmented Generation Models, Embeddings and LLM Performance Using Context-Aware Models for Clinical Pathways and Outcomes

Jakka Sairamesh, CapsicoHealth, Inc, PALO ALTO, CA, United States, Shloke Meresh, Laurent Hasson, Supraja Krovvidi, Siham Belgadi

RAGs (Retrieval Augmented Generation) have become instrumental in driving valuable, semantically-oriented, context-driven inputs to LLMs to reduce errors, hallucinations and unusual responses from LLMs . In this presentation, we will focus on healthcare oriented questions and responses using RAGs and compare multiple LLMs (including LLAMA2, LLAMA3, GPT4, Mistral and several other models) on their performance to predict future clinical pathways for a range of conditions such as cardiovascular, behavioral health and oncology-care. We found that depending on the healthcare focused contextual searches (including text and image data) and distance vector based models, RAGs can augment the LLM accuracy for specific scenarios in clinical decision-making. We found over 75% accuracy with LLAMA2 and LLAMA3 models when compared to 95% accuracy in using context-specific using the same data sets but augmented with semantic models. Our results have demonstrated that context-aware LLMs outperform publicly available LLMs with RAG driven inputs.

2 - Augmented LLMs Through Context-Aware Semantic Search and Automated detection of Hallucinations for Healthcare Decision-Tools

Supraja Krovvidi, CapsicoHealth, San Mateo, CA, United States, Shloke Meresh, Siham Belgadi, Sairamesh Jakka, Laurent Hasson

Abstract: As LLMs become more widely used, it's crucial to address potential risks and ensure high accuracy for decision-making, especially in healthcare. A major concern is 'hallucination', where LLM outputs diverge significantly from expected responses. Our study's main aim is to evaluate, detect and analyze continuously the degree of hallucinations in responses from custom-built LLMs on patient treatment pathways and clinical trials.

Hypothesis: Our hypothesis is that automating early-detection of deviations in LLM outputs can ensure continuous evaluation and reliability of LLMs during operational clinical use. We trained our custom-built LLM with trusted patient pathways and real-world clinical drug data

Methodology: Our approach involves leveraging Retrieval-Augmented Generation (RAG) architecture which enables the user to ask questions about a certain patient journey pathway or trends regarding a certain drug and potentially get a response considering both semantic similarity and context relevance. Our hallucination detection module is based on discrepancy detection between generated responses and retrieved data. This helps us in identifying significant deviations from the retrieved context, indicating potential hallucinations or inaccuracies. Techniques such as cosine similarity, semantic overlap metrics, or tree-based heuristics are employed for hallucination detection.

Evaluation: To enable trustworthiness for decision-making, several metrics are needed to ensure high accuracy and low hallucination rates. Metrics such as precision, recall, and F1 score are used to evaluate context-aware semantic search effectiveness and hallucination detection accuracy

3 - Could Provider and Patient Characteristics Influence the Outcomes of Provider Coaching in Emergency Care?

Sara Nourazari, California State University, Long Beach, Long Beach, CA, United States

This study examines the influence of patient and provider characteristics on emergency department (ED) quality scores and the effectiveness of professional coaching in improving these scores. We analyzed over one million ED visits, combining workflow data, patient feedback, and coaching records. Our results demonstrate that specific patient and provider demographics significantly impact the success of coaching interventions. This highlights the need for tailored coaching strategies to maximize effectiveness within the dynamic ED environment.

MC74

Regency - 705

Toward Resilient Power Systems: Advanced Methodologies for Uncertainty Management During Extreme Events

Invited Session

ENRE: Electricity

Chair: Yujia Li, lawrence berkeley national laboratory, 1 cyclotron rd ms 65a-110, El Cerrito, 94720, United States

Co-Chair: Yunhe Hou, The University of Hong Kong, Hong Kong, Hong Kong

1 - An Online Chance Constrained Optimization Framework in Resilient Microgrids Using Analytical Stability Sensitivity

Jun Wang, The university of Hong Kong, Hong Kong, Hong Kong

Offline optimization has traditionally been favored in power systems due to dependable generation and load forecasts. However, as renewable energy sources (RESs) and demand response technologies become more prevalent, the relevance of offline optimization is diminishing. This shift has prompted a greater focus on online optimization techniques in the power systems. To manage RESs uncertainties and stability issues in isolated resilient microgrids, this paper introduces a novel online chance constrained optimization framework accounting stability constraint. Thanks to proposed analytical sensitivities and linear reformulations, compared to conventional Monto-Carlo simulations, the computational burden is much lower, which can be acceptable to online optimization. Firstly, we propose a bi-level optimization problem, in which the upper level aims to minimize the expected generation cost without violating the stability chance constraint and the lower level gives the stability index in terms of a semi-definite program (SDP). Secondly, we apply the Gaussian mixture model (GMM) to handle the RES uncertainties and introduce analytical sensitivity analysis to reformulate chance constraints into linear deterministic versions. By incorporating linearized constraints, the bi-level tractable model can be solved by the Benders decomposition-based approach. Simulation results on the IEEE 33-bus microgrid test system reveal that compared to benchmarking approaches, the proposed model can converge in 1 minute and 30 times faster with more accurate solutions.

2 - Distribution System Restoration with Temporal Flexibility from Building AC Systems under Endogenous and Exogenous Uncertainties

Ruolan Cui, University of Hong Kong, Hong Kong Island, Hong Kong, Shunbo Lei

3 - Planning Resilient Distribution Grids

Miguel Heleno, Lawrence Berkeley National Laboratory, San Francisco, CA, United States, Alexandre Moreira

Distribution grid reliability and resilience have become critical issues for utilities and regulators, particularly due to the increasing severity of extreme events. Utilities are now considering significant investments in distribution grid assets to mitigate the impact of major outages. Effective quantification of the economic and risk-mitigation benefits of these investments to regulators is essential for the approval process. Current industry practices for reliability and resilience planning largely overlook these risk considerations and fail to properly quantify them. This talk discusses innovative optimization-based methodologies for distribution grid planning that balance the risk of outages cased by high impact low probability (HILP) events with the ones related to routine failures. This methodologies are illustrated with examples of reliability/resilience investments in real networks.

4 - Modelling Load Redistribution Attacks in Integrated Electricity-Gas Systems

Rongpeng Liu, McGill University, Montreal, QC, Canada, Xiaozhe Wang, Bo Zeng, Rawad Zgheib

As the proportion of gas keeps rising in electricity generation, integrated electricity-gas systems (IEGSs) have been deployed massively, facilitating gas transmission to gas-fired power generators and therefore mitigating generation costs. However, these deployments intensify the reliance of IEGSs on cyber infrastructures, such as communication systems, which in turn renders IEGSs progressively susceptible to cyberattacks. In this presentation, we delve into the paradigm and stealthy conditions of load redistribution (LR) attacks on IEGSs and propose a bilevel mixed-integer model to identify the most severe LR attack from an economic perspective. Under a mild assumption, we prove that the proposed model does not exclude any possible upper-level attack, which differs from existing models that may yield suboptimal solutions. A modified reformulation and decomposition (R&D) algorithm is developed to solve this model in a master-subproblem framework. Particularly, we design a new subproblem to address potential infeasibility issues in the master problem. Accordingly, two types of cuts are added to the master problem for ensuring algorithm feasibility and solution optimality. Testing results validate the effectiveness of the proposed model and solution method.

5 - Resilience Enhancement in Power Distribution Systems: Chance-Constrained Model with Decision-Dependent Atoms

Zeyu Wang, The George Washington University, Washington, DC, United States, Miguel Lejeune, Payman Dehghanian, Jinshun Su

Mobile power sources (MPSs) offer considerable promise to facilitate the exchange of spatiotemporal flexibility in the power distribution system (PDS) and enhance its resilience to extreme emergencies. The service restoration process can be made more efficient by strategically prepositioning MPSs in advance of and allocating them during high-impact low-probability extremes. This paper introduces a new decision-dependent chance-constrained model, where the atoms of the probability distribution of a random variable are distorted by decisions. The model is designed to address the service restoration problem in PDSs and optimize the pre-disaster deployment of MPSs. The proposed service restoration scheme is formulated as a mixed-integer nonlinear joint chance-constrained model with nonconvex continuous relaxation. We develop an efficient method to reformulate the chance-constrained model with decision-dependent uncertainty as an equivalent mixed-integer linear problem. Case studies demonstrate the effectiveness of the proposed restoration scheme in boosting PDS resilience and the significance of incorporating endogenous uncertainties in the service restoration process.

MC75

Regency - 706

Sustainable and Responsible Operations

Invited Session

ENRE: Environment and Sustainability

Chair: Can Zhang, Duke University, Durham, NC, United States

1 - When Where Watt: Harnessing the Value of TIME and Location of Electricity Generation for Renewables

Vishrut Rana, The Wharton School, Philadelphia, PA, United States, Christian Kaps, Serguei Netessine

Investments in renewable energy have surged worldwide in recent years, with over $1.7 trillion spent globally on clean energy investments. This surge can be attributed to growing policy support, energy security concerns, and most importantly, cost competitiveness of renewable energy. While technological advancements have ensured that costs for wind turbines and solar panels have plummeted, these lower costs do not translate to an increase in profits for renewable energy generation. Renewable power plants generate lower revenue compared to their conventional counterparts like coal and gas because renewable electricity is often generated in locations and at times with low wholesale market prices of electricity. To counter this, we propose a new site selection metric for renewable power plants – Quality Adjusted Power Value (QAPV) – which captures monetary value of electricity by accounting for time and location of renewable generation. We use geospatial satellite data to validate this metric on wind power plants installed in Texas between 2007-2014, and find that annual revenues for these plants can increase up to 30% if they were sited using QAPV. We conclude our discussion by identifying mechanisms that lead to sub-optimal locations being chosen for renewable power plants. Our work contributes two insights – for renewable developers, our site selection metric can be used to identify most profitable investment sites, and for policymakers, our results on counterfactual revenue can enable discussions around policies that streamline efficient investments in renewable energy.

2 - Buyer Management of a Supplier&Rsquo;S Social Responsibility Risk: Prevention, Detection, and Remediation

Junhao Yu, Miami University, Oxford, OH, United States, Tim Kraft

We study how retailers can proactively manage their suppliers' responsibility risk. We focus on a retailer's risk management strategy and consider three tactics the retailer can deploy—prevention, detection, and remediation—and study how combining the three can help the retailer achieve optimal results.

3 - When Workers Watch: the Impact of Metering Engagement on Energy Efficiency

Donggyu Jeon, Indiana University - Kelley School of Business, Bloomington, IN, United States, Owen Wu, Christopher Chen, Amrou Awaysheh

This study examines the impact of worker engagement with metering systems on energy efficiency in manufacturing settings. Metering systems, which form a vital part of the operational infrastructure, provide comprehensive data on energy and electricity usage, consumption flows and patterns, and CO2 emissions. Employing a rigorous empirical approach, this paper aims to empirically demonstrate the relationship between the dynamics of user-system interactions and energy efficiency outcomes. Our empirical findings reveal a positive relationship between the intensity of worker interaction with these systems—measured in terms of interaction frequency and duration—and the overall energy efficiency at the manufacturing facilities. In addition, qualitative aspects of interaction—specifically, the manner in which workers interact with the system and utilize or access different types of data—influence energy efficiency to varying extents. We consistently observe that heightened worker involvement with metering system and data leads to more energy-efficient operations at the manufacturing sites. Notably, we also find that user interactions focused on operational or energy-related tasks lead to greater energy efficiency compared to generic, administrative interactions. Overall, our findings show that enhanced access to information and broader data coverage, coupled with increased scrutiny of metering data by workers, translate into more informed decisions and practices that promote energy efficiency.

4 - The Value of Strategic Management of Residential Energy Storage Using a TIME-Invariant Control Policy

Na Rea Cho, Wichita State University, Wichita, KS, United States, Youngsoo Kim, Karthik Murali, Mesut Yavuz

While the deployment of renewable energy sources as an alternative to fossil fuels has garnered global attention, the intermittent and unpredictable nature of renewables has spurred a growing interest in energy storage. However, effective management of energy storage is far from trivial for residential consumers owing to various compensation and pricing structures, as well as uncertainties in electricity demand and generation.

In this paper, we establish an effective time-invariant policy to manage home energy storage and to improve the economic returns from residential solar and grid-tied battery systems. Through a numerical experiment, we show that this simple yet effective control policy substantially increases the value of home energy storage while performing nearly as well as battery optimization under certainty, thus leaving little room for improvement from more complex, forecast-driven approaches. From a policy perspective, our results indicate that imposing restrictions on the purchase of electricity from the grid to charge the battery is counter-productive to the adoption of this technology and the health of the grid and environment. In addition, through case studies, we show that strategically managing home energy storage makes this technology significantly more attractive even in communities where it would add no value when managed passively. Furthermore, real-world implementation in a community setting validates the effectiveness and robustness of the battery management policy established in this paper, highlighting its resilience to changes in billing and compensation plans. Our findings provide valuable insights into the effective utilization of residential energy storage systems under evolving regulatory and market conditions.

MC76

Regency - 707

Energy Systems Modeling and Optimization VII: Macroenergy Systems (MES)

Invited Session

ENRE: Other Energy

Chair: Raziye Aghapour, University of Texas at Arlington, ARLINGTON, TX, United States

1 - Market equilibria with agent-specific capital costs in energy system models – towards a consistent model formulation

Christoph Weber, University Duisburg-Essen, Essen, Germany

It is common practice to use linear optimization approaches with a single, uniform discount rate in energy system modelling. This is also consistent with the standard central planner interpretation of such models. At the same time, it is well known that outcomes of appropriately formulated optimization models may be interpreted as competitive market equilibria. Empirical evidence yet suggests that investors in energy markets face differentiated capital costs (e.g. Steffen 2020, Polzin et al. 2021) depending on technology, market and investor characteristics. There have been pragmatic approaches to incorporate such differentiated capital costs into energy system models (e.g. Loulou et al. 2016, Tash et al. 2019, Lonergan et al. 2023). Through a thorough analysis of KKT conditions, a theoretically consistent solution may yet be established with succinct and explainable economic features. We illustrate the approach with a stylized application to capacity expansion in Germany and discuss how the outcomes may be linked to standard results. Also the implications for energy and climate policy are highlighted.

2 - Exploring Transmission Solutions for Texas' Renewable Energy Surge

Raziye Aghapour, University of Texas at Arlington, Arlington, TX, United States

Texas, a significant energy producer, has experienced a notable surge in the integration of renewable energy lately, with expectations of continued growth, particularly in the renewable-rich region of West Texas. However, this region's distance from major demand centers in the eastern cities, coupled with the lagging expansion of Texas's transmission infrastructure, poses challenges to accommodating this growing renewable energy production. These obstacles could lead to increased congestion and constraints within the ERCOT system. In this study, we aim to address these challenges by analyzing potential transmission enhancements to alleviate export constraints from West Texas. Using the Regional Energy Deployment System (ReEDS), a macro energy system modeling tool, we will explore various scenarios. Firstly, we will evaluate the effectiveness of ERCOT's transmission planning in mitigating congestion, followed by a comparison with our own findings, highlighting the respective strengths and weaknesses of each approach.

3 - Effects of Storage Participation on Wholesale Electricity Markets

Zhenhua Zhang , UCSD, San Diego, CA, United States

The strategic behaviors of storage systems that participate in wholesale electricity markets can influence market prices, system costs, and profit distributions among generation types. We develop a bi-level model to examine how these impacts vary across participation requirements in the Western Electricity Coordinating Council region. The upper level includes a storage profit maximization model and the lower level represents the market operator's economic dispatch problem. We find that strategic behaviors can significantly increase storage profit but increase system cost only by a small margin under various renewable penetration levels relative to the case when storage is ISO-controlled. Market prices decrease during charging events but remain similar during discharging events, subject to electricity demand conditions and discharge offer prices. For storage that also participates in resource adequacy programs with minimum state-of-charge requirements, a low non-compliance penalty can help recover storage investments without significantly increasing system costs. Future designs on wholesale participation models and resource adequacy programs should consider how ISO requirements and storage strategic behaviors interact and evaluate the trade-offs between generator profitability and system costs.

4 - Solar adoption in MA across sectors

Zana Cranmer, Bentley University, Waltham, MA, United States

Solar energy is a key component of strategies to reach net zero emissions. Understanding the current drivers of adoption of solar energy, particularly photovoltaics, is a key element to developing policies to address inequities and gaps in access. This work looks across sectors, including residential and commercial solar projects. Beyond demographic factors like income and homeownership, do measures of environmental attitudes and solar-focused municipal programs increase or accelerate adoption? What are the implications for reaching state targets for renewable energy, emissions, and equity? The identified drivers will then be applied to future projections for energy use and adoption.

5 - Optimizing Suburban Mobility: Scheduling Shared Autonomous Electric Vehicle Fleet with Stochastic Renewable Energy and Grid Load Variability

Luce Brotcorne, INRIA, Lille, France

The rapid development of Shared Autonomous Electric Vehicle (SAEV) fleets offers significant opportunities for enhancing suburban mobility. This paper addresses the challenge of integrating AEV routing with charging infrastructure, focusing on the stochastic nature of day-ahead renewable energy availability and grid load variations. Our study leverages scenario-based stochastic optimization to model and manage uncertainties. We propose a framework that combines AEV charging scheduling and passenger travel demand fulfillment in suburban areas equipped with solar PV-integrated charging stations. The primary objective is to minimize total operational costs, including travel and charging expenses, while penalizing any unmet passenger requests to ensure that the majority of passenger travel demands are met within their specified time windows. The mixed-integer linear programming (MILP) model incorporates constraints such as vehicle battery capacities, time windows for passenger trips and charging station availability. Multiple scenarios are generated based on historical data, each representing potential solar PV output and grid load realizations for the upcoming day. The MILP is solved for each scenario and the solutions are used to formulate strategies that optimize the system under uncertainty. A case study of a suburban French low voltage (LV) grid and transportation system demonstrates the effectiveness of our approach in reducing costs and enhancing system reliability. Our model reduces peak grid demand and costs associated with charging the SAEV fleet by aligning vehicle charging times with periods of high solar PV output.

MC77

Regency - 708

Data-Driven Optimization for Reliable Machine Learning (II)

Invited Session

Computing Society

Chair: Vasileios Digalakis, HEC Paris, Jouy-en-Josas, France

Co-Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

1 - A Mathematical Optimization Model for Solving the Contextual Multiobjective Inverse Ideal Point Problem

Nuria Gómez-Vargas, Universidad de Sevilla, Sevilla, Spain, Rafael Blanquero, Emilio Carrizosa

We propose a Mathematical Optimization model tailored to address the challenging Contextual Multiobjective Inverse Ideal Point Problem (CMIIPP). The CMIIPP involves yielding the underlying objective functions of a multicriteria decision-making model based solely on a set of pairs of concurrently observed context and decisions, particularly focusing on the ideal point scenario in multiobjective optimization. By formulating the CMIIPP as an optimization problem, our approach offers a framework to uncover the implicit preferences governing the decision-making process that led to a given ideal point solution preserving its consistency. Furthermore, considerations are made for incorporating domain-specific knowledge to enhance applicability and exploring other requirements such as interpretability.

2 - Global Optimization: a Machine Learning Approach

George Margaritis, Massachusetts Institute of Technology (MIT), Cambridge, MA, United States, Dimitris Bertsimas

Many approaches for addressing Global Optimization problems typically rely on relaxations of nonlinear constraints over specific mathematical primitives. This is restricting in applications with constraints that are black-box, implicit or consist of more general primitives. Trying to address such limitations, Bertsimas and Ozturk (2023) proposed OCTHaGOn as a way of solving very general global optimization problems by approximating the nonlinear constraints using hyperplane-based Decision-Trees and then using those trees to construct a unified mixed integer optimization (MIO) approximation of the original problem. We provide extensions to this approach, by (i) approximating the original problem using other MIO-representable ML models besides Decision Trees, such as Gradient Boosted Trees, Multi Layer Perceptrons and Suport Vector Machines, (ii) proposing adaptive sampling procedures for more accurate machine learning-based constraint approximations, (iii) utilizing robust optimization to account for the uncertainty of the sample-dependent training of the ML models, and (iv) leveraging a family of relaxations to address the infeasibilities of the final MIO approximation. We then test the enhanced framework in 81 Global Optimization instances. We show improvements in solution feasibility and optimality in the majority of instances. We also compare against BARON, showing improved optimality gaps or solution times in 11 instances.

3 - The Art of Selection: Large-Scale Optimal Experimental Design

Baizhi Song, London Business School, London, United Kingdom, Jean Pauphilet

Despite the recent explosion in data availability, data collection in certain fields remains challenging and expensive. Consequently, it is crucial to collect a subset of data samples that retains as much information as possible from the full population. This issue aligns with a fundamental problem in statistics: optimal experimental design. Recent examples of this problem in environmental sciences and agriculture introduce new challenges: (i) the total size is in the order of 10,000, (ii) data collection (e.g., via drones) is subject to complex operational constraints.

In this paper, we propose a discrete optimization approach that addresses both challenges. Our approach first solves the relaxation via outer approximation. On synthetic datasets, our approach extends the solvable problem scale by an order of magnitude (from 3,000 to 30,000). The corresponding Mixed Integer Optimization (MIO) problem benefits from a warm start of the solution and valid cuts generated during the relaxation phase. Our approach can include various geospatial/route constraints. In the context of ocean cleaning applications, we further investigate the trade-off between information and plastic collection.

4 - An online assignment and routing problem for the logistics of organ transplants

Sean Lo, MIT Operations Research Center, Cambridge, MA, United States, Alexandre Jacquillat

TransMedics is a US-based healthcare startup which operate devices for organ transplants. These portable devices preserve the shelf-life of donated organs outside the donor's body, allowing for less time-dependent organ donor-recipient matchings and improved healthcare outcomes. We help Transmedics solve a complex logistics problem on a rolling horizon with a hybrid assignment / routing structure: we assign surgeons, technicians, and devices to organ donors, and route them to organ recipients via a fleet of planes. We propose an online algorithm which takes into account uncertainty in case arrivals, and demonstrate how it outperforms a greedy approach. We deploy our method in TransMedics' operations center and show that we significantly improve on the volume of transplants served in real life.

MC78

Regency - 709

Advanced Optimization for Mixed Integer Programming

Invited Session

Computing Society

Chair: Mikhail Bragin, University of California, Riverside, Marina del Rey, CA, United States

1 - Enhancing Quantum Optimization Scalability: A Singular Transformation Approach to Unit Commitment

Peng Zhang, Stony Brook University, Stony Brook, NY, United States, Mikhail Bragin, Yifan Zhou

Resilience-constrained unit commitment (RCUC) is a quintessential strategy for meeting electricity demands and boosting grid resilience against extreme events. However, solving RCUC is a long standing challenge due to its mixed-integer, nonlinear, and non-convex nature. The recent successes in exploiting the potential of quantum advantages spread hope to the power industry for achieving a ‘quantum leap’ of grid analytics. Our experiments have proven the feasibility of building quantum UC algorithms such as quantum alternating direction method of multipliers and quantum surrogate Lagrangian relaxation. However, these quantum algorithms are difficult to scale, restricted by the limited number of qubits, low connectivity between qubits, and coherent and incoherent errors, etc. In this talk, we will describe a promising method for tackling practical RCUC problems by empowering the decomposition and coordination-supported framework with singular transformation. This singularity-based method will result in a highly expressive quantum circuit with manageable quantum operators. The new method is expected to achieve a complexity independent of the system dimension, which will pave the way for the utilization of limited quantum resources to potentially solve the large-scale mix integer optimization problems. We will demonstrate the quantum RCUC algorithm's accuracy, efficiency, and resilience on representative power systems with NISQ devices.

2 - Acceleration of Level Adjustment for the Polyak Stepsize: Applications to Mixed Integer Programming

Mikhail Bragin, University of California, Riverside, Riverside, CA, United States, Anbang Liu, Zuzhao Ye, Nanpeng Yu, Osten Anderson

The Polyak stepsize has been widely used for convex problems. However, the stepsize requires a generally unknown optimal value. To guarantee convergence, we develop a novel level adjustment approach to successively estimate the optimal value with provable convergence. This is achieved by developing a decision-based procedure through a novel easy-to-solve "Polyak Stepsize Violation Detector" (PSVD) linear constraint satisfaction problem. Once a violation is detected, a tighter level value is obtained. Furthermore, because subgradients may be computationally costly, we develop an approximate subgradient method. Through a series of empirical tests of convex functions with diverse characteristics, we illustrate the practical advantages of our approach as compared to existing methods. For Mixed-Integer Programming (MIP) problems, our integration of Lagrangian Relaxation with the Polyak stepsize demonstrates significant efficiency gains, optimizing Lagrangian dual problems in operations research, power systems, transportation, and manufacturing sectors leading to speedups of 2-3 orders of magnitude as compared to standard MIP methods.

3 - Accelerating Branch-and-Price via Template Pricing

Luke Marshall, Microsoft Research, Redmond, WA, United States, Prachi Shah, Santanu Dey

Column generation (CG) for discrete optimization (i.e., branch-and-price), is a powerful method for solving large-scale and highly complex problems. However, it also faces challenges with convergence when degeneracy occurs. Many methods have been explored in the literature to stabilize dual values, reduce degeneracy, or generate compatible columns, with varying success. As an alternative, we propose a new *template pricing* approach, which modifies the standard subproblem that generates new columns. It uses additional objectives and constraints to guide similarity between new columns and a dynamic set of templates.

Template pricing aims to achieve three goals: (i) generate columns that are not only profitable, but also diverse-yet-compatible with other columns; (ii) directly avoid problems with dual instability to speed up the convergence of the CG algorithm; and (iii) can be implemented easily and efficiently. We show that template pricing can be applied to various types of mixed-integer programming (MIP) problems, and present different ways to select suitable templates. Moreover, we also show that our approach is compatible with standard dual-value stabilization techniques for even greater impact.

We perform extensive computational experiments on a variety of benchmark instances from the literature and compare our approach with other methods. Our results show that template pricing can significantly accelerate the CG algorithm, with some instances solving 500x faster. To the best of our knowledge, this is the first work that proposes and evaluates template pricing as a general acceleration method for column generation in integer programming.

4 - Using Lagrangian decomposition to coordinate order release planning and production scheduling

Hubert Missbauer, University of Innsbruck, Innsbruck, Austria, Reha Uzsoy, Gregor Blossey

Optimization models that determine order releases for a multi-period planning horizon simultaneously perform production smoothing and cycle time control by controlling the level of work in process inventory in the manufacturing system. The crucial modelling challenge is that of anticipating the effects of a specified order release schedule on the time-dependent cycle time distribution.
Previous authors have sought to improve this anticipation by decomposing the order release problem into a production smoothing subproblem, usually represented as a linear program, that determines the order releases for given time-dependent lead times, and a cycle time estimation subproblem for given releases, usually addressed by discrete-event simulation. The subproblems are solved iteratively until convergence. However, this procedure can fail to converge, and it has been shown recently that it is a flawed application of mathematical decomposition.

We propose a Lagrangian decomposition approach that decomposes the order release problem into order release and production scheduling subproblems. Coordination is achieved through Lagrange multipliers that act as prices for capacities and penalize long lead times. We show that linear prices cannot coordinate the subproblems. We test a variant of augmented Lagrangian relaxation that ensures separability and the Alternating Direction Method of Multipliers (ADMM). Preliminary computational results show that both methods work in principle but show shortcomings in their convergence behavior. We discuss these open questions and outline alternative ways to apply Lagrangian techniques to coordinate the order release and the scheduling level. This also aims at contributing to the integration of mathematical programming and discrete-event simulation.

Monday PC

Flex C

Monday Poster Competition

Poster Session

Poster

Chair: Sripad Devalkar, Indian School of Business, Hyderabad, India

Co-Chair: Masha Shunko, University of Washington, Seattle, WA, United States

Co-Chair: Shengfan Zhang, University of Arkansas, Fayetteville, AR, United States

1 - Life After Warming: How Can We Make It Better Using Systems Engineering for Societal Good?

Buket Cilali, University of Oklahoma, Norman, OK, United States, Kash Barker, Andres Gonzalez, Ahti Salo

Forced displacement is a pressing global issue, requiring proactive international solutions for relocation and integration planning. Particularly in climate-driven cases, where slow-onset climate change impacts render regions uninhabitable for the upcoming decades, addressing this large-scale displacement in the long term requires a comprehensive approach. To this end, we propose a two-stage stochastic optimization model that analyzes diverse demand scenarios and societal impacts to determine where and when people should go in order to assist decision-makers in preparing for future movements effectively.

2 - Bringing Sustainability to Mining Regions with an Evolutionary Interactive Multi-Objective Optimization with Multiple Stakeholders

Gen LI, Technical University of Munich, School of Management, Center for Energy Markets, Munich, Germany, Svetlana Ikonnikova

The rising demand for critical materials, driven by global decarbonization, is intensifying mining operations in the Global South. This trend necessitates research on sustainable mining practices that balance economic viability with social responsibility and environmental sustainability. Despite companies' focus on mutual financial gains, stakeholders including local communities, regulators, and downstream players frequently voice concerns over the neglect of their conflicting interests. This research aims to develop a practical AI-based simulation framework for (1) assessing the impact and outcomes associated with mining and (2) finding a compromise solution determining the mining setup and operations, acceptable to the affected stakeholders.

Built on interactive multi-objective optimization (iMOO), our framework is enhanced by integrating evolutionary algorithms with models of multilateral bargaining and choice behavior. This integration takes into account the evolving preferences and relationships of stakeholders, aspects often overlooked by traditional iMOO. The hybrid model effectively manages conflicting priorities and interests, identifies equitable solutions, and ensures that the benefits of mining developments are distributed fairly and responsibly. Additionally, the framework incorporates multi-actor multi-criteria analysis and utilizes advanced interactive visualization techniques to help stakeholders understand and evaluate the trade-offs among objectives. This aids in prioritizing final solutions and building consensus on sustainable practices.

Furthermore, our scalable simulation of the decision-making process demonstrates how stakeholders can learn about and negotiate trade-offs, thereby reaching a consensus on sustainable mining practices. This project contributes significantly to the field of operations research by illustrating how sophisticated analytical tools can facilitate sustainable decision-making in complex, multi-stakeholder environments.

3 - Integrating Satellite Imagery to predict and verify Supply Chain Carbon Emissions

Alexander Rose, WHU, Vallendar, Germany, Stefan Spinler

Accurate supply chain carbon emissions data is needed for successful decarbonization. However, data lacks accuracy with limited access to supplier primary information and heterogeneous calculation standards. We apply machine learning algorithms to compare and integrate satellite emissions data with public disclosures to predict and verify suppliers' emissions with increased accuracy.

4 - Advanced Satellite Scheduling Strategy for Snow Water Equivalent (SWE) Monitoring

Hadis Banafsheh, Arizona State University, Tempe, AZ, United States, Paul Grogan

Accurate monitoring and forecasting of Snow Water Equivalent (SWE) are essential for effective water resource management. Space-based observational data are sourced from two Earth Observation Satellite (EOS) types: publicly-available satellites and fee-based taskable satellites, which have unpredictable and limited availability.

This research aims to enhance the scientific value of satellite observations by implementing reward-based optimization strategies that determine the optimal timing and selection from both satellite types in the western United States. Our methodology utilizes shapefiles to define specific regions of interest and employs the SNODAS data set, a modeling and data assimilation system developed by NOHRSC1, to establish a baseline for SWE. This methodology enables the optimization of observations by calculating satellite ground tracks over designated areas, improving the effectiveness of observations to infer the true environmental state under budget limitations.

We have developed an analytical evaluation platform to compare various satellite scheduling strategies. Our study includes three scenarios: using publicly-available satellites alone, combining publicly-available and taskable satellites with an unlimited budget, and employing both satellite types under limited budget conditions with static and dynamic scheduling. Our experiment applies these strategies to satellite observations for an entire snow season. By analyzing resource utilization and observation frequency across these scenarios, we aim to demonstrate measurable improvements in the effectiveness of satellite observation scheduling. This comprehensive approach both tests theoretical models and provides practical insights into their operational viability, advancing satellite resource management and environmental forecasting practices.

5 - An Application of Multi-Objective Optimization in Green Supply Chain Network Design

Labiba Noshin Asha, University of Arkansas, Fayetteville, AR, United States, Harun Pirim, Nita Yodo, Zaidur Rahman

Purpose: This study proposes a multi-objective model to balance costs and carbon emissions in green supply chain management, aiming to simultaneously minimize total costs and carbon emissions across the entire supply chain network.

Originality:

•The multi-objective model is proposed by integrating various harvesting styles for green supply chain network design.

•A heuristic solution method is introduced, enhancing the model's applicability for both small and large-scale problems.

Approach: The study examines a supply network between crop production zones and processing plants. Using the weighted sum method, the multi-objective model is simplified. The heuristic approach effectively evaluates potential solutions, providing a practical model to enhance decision-making in green supply chain management.

Findings: The optimal solution is determined for a small-scale problem and subsequently compared with the heuristic solution. The result indicates a close proximity between the optimal and heuristic solutions for small-scale problems. Moreover, the result also shows that in more complex problem scenarios where finding the optimal solution is challenging, this heuristic approach effectively offers a practical and viable solution.

6 - DSS for Climate Mitigation Using Indifference Probabilities to Measure Impact Utilities

Kunsinee Srivichaiin, Boston University, Boston, MA, United States, Yu Gao, Kaumudi Dande

This poster focuses on the development of a decision support system to mitigate climate change impacts on a university campus. The project was funded by the Boston University (BU) Institute for Global Sustainability. The system helps a facility manager choose the best alternative among 3-5 retrofitting options. The challenge of quantifying climate risk impact is overcome by applying indifference analysis, which accounts for disruptions to operational activities and associated ripple effects. The R-Shiny web application requires a user to input the probability of the climate event, a list of mitigation options with their implementation costs, and the likelihood that the mitigation option will prevent the climate event from impacting the facility. Color-coded matrices are used to display uncertainty for indifference analyses by presenting sets of pairwise comparisons in an iterative manner using a binary search. The pairwise differences in the expected values of risk mitigation options, calculated from the indifference probabilities, are converted into option rankings utilizing the AHP methodology. The application is illustrated using a robotics laboratory facility located in a high wind risk area of the BU campus. Four risk response options were compared, including installation of windows with different design pressure rating and accepting the risk of window failure.

7 - Adapting the Inventory Routing Problem to Improve Water Delivery Scheduling for Water Insecure Households

Emerald Dudzinski, Virginia Tech, Blacksburg, VA, United States, Kimberly Ellis

Access to clean drinking water can be a luxury in impoverished communities in the United States. Nationally, more than 2 million people lack access to safe drinking water and indoor plumbing, and this figure does not include homes with a working tap of unsafe drinking water. Therefore, a far larger number of Americans face water insecurity. A recent examination of US Census and Safe Drinking Water Act violation data reveals that rural low-income and minority communities are significantly more likely to be burdened with unavailable or unsafe in-home drinking water. This research explores an alternative to traditional water infrastructure: how to optimally deliver drinking water to households without access to reliable, clean water sources by adapting the inventory routing problem (IRP) and creating heuristic algorithms to schedule efficient water delivery. The IRP and heuristics inform three key decisions: when to serve a customer, how much to deliver, and which delivery routes to use.

8 - Combining Fuzzy P-graph and Multi-Criteria Decision Making Method in the Case of Cogeneration System

TZUI TSAI, TUNGHAI UNIVERSITY, Taichung City, Taiwan, YI-HSIANG CHEN, TING YU LIN, Kuo-Ping Lin

With the increase in population and economic expansion, greenhouse gas emissions have become one of the significant factors contributing to the worsening phenomenon of global warming. Global natural resource use has already surpassed the Earth's renewable capacity. Developing cogeneration systems may be a potential solution to address the issues of energy depletion and carbon emissions. The P-graph has been widely applied to optimizing the structure of cogeneration systems. This study develops a new fuzzy P-graph method under real-world uncertainties and uses multi-criteria decision-making methods to consider the critical factors influencing cogeneration. Among the many feasible solutions, the optimal system structure is identified. Through the verification of a real-world case, it is found that the study results can provide decision-makers with comprehensive, feasible outcomes, avoiding subjective decisions by individual decision-makers regarding the system architecture. The fuzzy P-graph can handle data uncertainties, making the results more realistic.

9 - Win-win or Lose-win? Economic-Climatic Synergies and Trade-offs in Cattle Farming

Mario Guajardo, NHH Norwegian School of Economics, Bergen, Norway, Rasmus Bang, Ola Flaten, Julio Goez, Bjørn-Gunnar Hansen, Hanne Møller, Stine Samsonstuen, Jon-Kristian Sommerseth

Governments and societies strive to balance food production with environmental goals. In this context, identifying farm-level economic and climatic win-win and lose-win scenarios is essential, not only for farmers but also for policymakers and the broader society. This study aims to examine economic performance and greenhouse gas (GHG) emission intensity in combined milk and beef production on dairy farms. We present a quantitative framework to economically optimize production at farm level under various operational constraints and to calculate the corresponding GHG emissions. We apply this framework in Norway. Our results demonstrate that the effectiveness of various measures to increase economic performance and to reduce GHG emissions strongly depends on operational conditions. Moreover, they show that measures that produce win-win outcomes in some situations may result in lose-win or win-lose outcomes in others.

10 - Optimizing Vaccine Stockpiles: A Stochastic Model for Combination and Monovalent Vaccines

Ming Zhu, Rochester Institute of Technology, Rochester, NY, United States, Sheetal Aher, Ruben Proano

Vaccine shortages are one of the primary healthcare problems around the globe. Despite multiple initiatives to reduce vaccine shortages, supply interruptions are unexpected and unavoidable. To mitigate the risk of such interruptions, vaccine stockpiles have traditionally been the tool of choice. Vaccines that offer protection against more than one disease in one injection are. However, vaccine inventory models determine stockpile levels for a single product, ignoring the available supply and inventory levels of other vaccines which could partially satisfy the product's demands. This study aims to determine optimal stockpile levels for combination vaccines, considering the existing inventory of other vaccines offering subsets of their antigen content. We propose a stochastic tractable safety stock model that considers the availability of a pool of monovalent and combination vaccines in the stockpile that together can offer the same set of antigens as a vaccine that could face shortages. We contrast the recommendations from the stochastic safety stock model with a simulation and establish that the model recommendations are feasible and can effectively reduce the safety stock level, which is unique.

11 - A Quantum Optimization Algorithm for Electric Vehicle Charging Station Placement for Intercity Trips

Tina Radvand, University of Illinois at Urbana- Champaign, Urbana, IL, United States, Alireza Talebpour

Electric Vehicles (EVs) are set to revolutionize transportation by offering sustainable solutions to environmental challenges and the global energy crisis. However, their widespread adoption is hindered by inadequate charging infrastructure, which limits driving range and user convenience. Identifying optimal charging station locations in large transportation networks with minimal stations presents a well-known NP-hard combinatorial optimization problem, as the search space grows exponentially with the number of potential charging station locations.This paper introduces a quantum search-based optimization algorithm designed to enhance the efficiency of solving this NP-hard problem for both corridors and transportation networks. By leveraging parallel computation, amplitude amplification, and quantum phase estimation as a subroutine, the optimal solution is found quadratically faster than classical exact methods like branch and bound. The detailed design and complexity of a resource-efficient quantum circuit are discussed. The proposed quantum-optimization technique can also be applied to trip planning for EV drivers, providing guidance on optimal charging locations.

Monday Poster

Flex C

Monday Poster Session

Poster Session

Poster

Chair: Shengfan Zhang, University of Arkansas, Fayetteville, AR, United States

Co-Chair: Masha Shunko, University of Washington, Seattle, WA, United States

Co-Chair: Sripad Devalkar, Indian School of Business, Hyderabad, India

1 - Comparing Reformulation and Cutting Plane Methods in Solving Distributionally Robust Problems

Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States, Bo Zeng

This study extends the well-studied reformulation and cutting-plane methods in robust optimization (RO) to distributionally robust optimization (DRO) and chance-constrained problems (DRCCP). We test these approaches on moment inequalities and Wasserstein ambiguity sets, comparing their effectiveness and efficiency in handling these distributionally robust problems. By conducting extensive comparisons, our research aims to provide a comprehensive analysis of their practical applications in uncertain environments.

2 - A Deep Variational Bayesian Approach for Modeling and Condition Monitoring of Dynamic Systems

Nadia Tanzeem, University of Miami, South Miami, FL, United States, Ramin Moghaddass

This paper introduces a Deep State-Space modeling framework tailored for monitoring complex and dynamic systems operating under varying conditions. By integrating Recurrent Neural Networks (RNNs) with generative models in a variational autoencoder form, the proposed framework effectively approximates the complex behaviors inherent in systems with latent states. Emphasizing the enhancement of system reliability with restrictive assumptions on dynamic behavior, the model excels at discerning stochastic patterns observed in both dynamic measurement data and latent states within complex degrading systems. The proposed model can capture complex temporal dependencies through RNNs and deploy advanced probabilistic techniques like ancestor sampling for robust inference. Moreover, its versatility extends to accommodating both continuous and discrete latent states, thereby enriching the representation of underlying data dynamics. Primarily, the model finds utility in latent state estimation and Time of Event prediction, critical in predictive maintenance and reliability engineering contexts. We have tested this framework using both simulations and real-world data. Also, a case study utilizing a Wind Turbine dataset further substantiates the proposed framework's effectiveness in early fault detection, thereby showcasing its broad applicability across diverse industrial settings.

3 - Designing algorithms for autonomous mobile robotic systems in package sortation centers

Yujie Sun, Singpore Management University, Singapore, Singapore

We study an autonomous mobile robotic system in the context of packages sortation center with multiple storage pods and processing stations, operating within a finite multi-period horizon. The arriving packages are interconnected, each associated with a specific delivery route. At the start of each period, our task involves selecting the pods for stowing the arriving packages at each station. At the end of each period, certain routes are called for delivery according to the delivery plan, requiring the selection of pods for retrieving packages associated with these routes. Our objective is to minimize the total travel cost of the autonomous mobile robots. We formulate an integrated packages stowage and retrieval problem over multiple periods with both deterministic and uncertain arriving packages. To make the problem tractable, we propose two approaches. In the first approach, we adopt the idea of decomposition by categorizing storage areas into several storage classes, each containing multiple pods. We first choose storage classes to stow or retrieve the packages in each period, then we select pods from the chosen storage classes for stowing and retrieving packages. In the second approach, we adjust the decision-making process by initially stowing packages to pods or retrieving them from pods without specifying the processing station. Subsequently, we determine the stations responsible for handling the stowage and retrieval. Both approaches not only yield approximate optimal solutions but also significantly reduce computational costs.

4 - Comparing the performance of the collaborative market and segmented market of ride-sharing system: A dynamic maximal weighted matching algorithm

xin dong, Pennsylvania State University, State College, PA, United States, Tingxue Gu, Jose Ventura, Vikash Gayah

Ride-sharing platforms, such as Uber and Lyft, have transformed urban transportation by reducing total travel distances and lifting vehicle occupancy rate. However, the efficiency of these platforms is often limited by market segmentation, where driver-rider matches are confined to specific platforms, potentially underestimating optimal pairings. This study employs a dynamic maximal matching algorithm to compare the efficiency of a segmented (duopolistic) market structure with that of a collaborative market. We model riders and drivers as graph vertices, with addition and deletion of dynamic weighted edges representing travel distances. The graph (ride-sharing market) is partitioned into sub-graphs corresponding to each platform and employs the dynamic maximal weighted matching algorithm to find the maximal-matching assignment within each sub-graph. Our findings demonstrate that a collaborative market significantly enhances the match and share rates within the ride-sharing system by overcoming the barriers imposed by market segmentation. Additionally, we explore how the computational performance of the dynamic maximal weighted matching algorithm varies with changes in the number of nodes and edges. These results underline the potential benefits of cross-platform collaboration in the ride-sharing industry, encouraging entities to cooperate in a competitive environment to optimize the operational efficiency.

5 - Adaptive Experimental Design: Machine Learning Directed Optimization of Additive Manufacturing for Thermoelectric Materials and Devices

Ke Wang, University of Notre Dame, Notre Dame, IN, United States, Alexander Dowling

Solid-state thermoelectric (TE) generators (TEGs) are promising for powering distributed sensor networks, biomedical devices, wearable electronics, and beyond. However, the efficiency of TEGs is limited by TE materials and their manufacturing methods. The Edisonian search, widely adopted by experimentalists to improve TE materials performance, remains expensive, slow, and unsuitable to handle large search space, e.g., $O(10^6)$. In recent years, artificial intelligence and machine learning have offered new approaches to leverage data to overcome these challenges.

This poster will focus on the data-driven optimization framework, integrating first-principle, machine learning, and Bayesian optimization to reduce the order of magnitude of fewer experiments than the Edisonian search. The application of the proposed framework for the optimization of TE materials and corresponding manufacturing methods, including material composition, 3D printing, and sintering, will be highlighted.

6 - A GNN-accelerated Heuristic for Solving Boolean Quadratic Programs

Natalie Isenberg, Pacific Northwest National Laboratory, Richland, WA, United States, Weimin Huang, Jan Drgona, Bistra Dilkina, Draguna Vrabie

Boolean Quadratic Programs (BQPs) are widely applicable in various applied domains, such as facility location and wind farm layout planning. However, solving large-scale instances of BQPs using traditional exact methods often becomes computationally impractical due to combinatoric complexity and nonconvexity. To address this, we propose a machine learning-based adaptation of the Undercover heuristic tailored for BQPs. This approach leverages the insight that fixing a subset of variables, specifically the minimum vertex cover of the Hessian graph, can simplify the problem by linearizing it, thus making it more tractable. While identifying this cover set is computationally efficient with existing algorithms, our novel contribution lies in integrating a graph neural network (GNN) within the heuristic to predict high-quality solutions for the cover variables. During inference, the trained GNN model is used to assign values to the cover variables, which are then fixed, reducing the BQP to a simpler linear sub-problem that can be solved efficiently. This method is particularly motivated by the need to solve the optimal wind farm layout problem, which can be formulated as a large-scale cardinality constrained BQP.

7 - Transistor routing algorithm in analog circuit for memory semiconductor design automation

SunJin Han, POSTECH, Pohang, Korea, Republic of, Beomhi Lee, Hyungjun Park, Byung-In Kim, Dong Gu Choi, Joonrak Kim, Hyunkyu Kim, Youjung Lee

In semiconductor manufacturing companies, the analog circuit design has been performed manually, so it is a time-consuming and error-prone task. Recently, the research on the automation of the task has been conducted actively. This study focuses on the automation of a design task that wants to minimize process variations while adhering to Common Centroid constraints and considering the importance of various evaluation metrics simultaneously. We develop a method for the task in which a rule-based approach derives diverse transistor placement results and utilizes a genetic algorithm to produce various routing results corresponding to each placement. To increase computational efficiency, the genetic algorithm initially evaluates the lower bound of fitness efficiently through analytical function. Subsequently, it selects elite sets for employing commercial solvers to derive accurate routing results. The validity of the developed algorithm is verified through test cases formulated in collaboration with industry experts.

8 - Control Policies for Telehealth Connecting Patients to Doctors

Shuwen Lu, Cornell University, Ithaca, NY, United States, Jamol Pender, Mark Lewis

This paper considers a sequential decision-making scenario in minute clinics faced by nurse practitioners (NPs) who accompany each patient through their visits. After an initial diagnosis of the patients, NPs have a decision to make on whether to continue the treatment themselves or to seek a collaborative telemedicine service with a (dedicated) general physician (GP). A Markov Decision Process (MDP) clearing system model is constructed and we are interested in a control policy that minimizes the cost of returning to the normative state after a demand spike.

Structural properties of the optimal policy have been investigated under different combinations of system parameters. The heuristics that are easily implementable given the parameter inputs are designed as approximate policies based on the observation and the proofs. The proposed heuristic policies have demonstrated their performance by showing only around 0.1\% worse than the optimal on average and outperforming the baseline policies. Finally, the insights are provided to both the decision-makers and the system designers.

9 - The Individual’s Perception of AI and Its Effects on the Organization’s AI Initiatives

Catherine Petersen, Johnson & Wales University, Mont Vernon, NH, United States

Investment in Artificial Intelligence (AI) does not result in returns for business despite improving technology. This study investigated one potential reason for the failure of AI when implemented within an organization with the intent to drive success with this technology. AI is unique amongst technologies due to its ability to inspire emotions from potential users and these emotions can affect the individual’s willingness to adopt the technology.

Using an explanatory sequential research design, this study examined the relationship between how an individual perceives AI and their willingness to accept the technology in their role within an organization. It also considered the role of experience and an individual’s perception of AI.

The results found that including AI in the description of the initiative decreased support from 81.5% to 54.8%. There was also a significant relationship between an individual’s perception of AI and their support of an initiative that specifically names AI ($p < .05$). In interviews, nine out of ten respondents discussed job replacement and showed a tension between hopes and fears with regard to the technology. No statistically significant relationship was found between an individual’s experience with AI and their perception of AI but there was a relationship between experience and their expectations for effort and performance. Interviews showed that an individual’s perception of job replacement or challenges with the technology were not affected or increased with experience using AI. Based on these findings, managers can better determine if their organizations are ready to implement AI successfully.

10 - Multiple Objective, Shortest Path-Based Running Route Optimization

John Sooter, University of Arkansas, Fayetteville, AR, United States

Runners prefer routes of an exact distance with desirable characteristics such as elevation change, pedestrian-friendly roads, and a natural setting. Finding routes that match these characteristics is important for the safety and enjoyment of a distance running routine. By capturing these attributes using OpenStreetMap data, one-to-all shortest path-based optimization can be used to generate high-quality routes as an alternative to finding routes by hand or running the same routes repeatedly. Both single and multiple objective variants of Dijkstra’s algorithm are explored to generate out-and-back running routes. Selecting the candidate destination nodes based on a radial distance criterion, we can guarantee the lower bound of feasible solutions to be the shortest path distance, then trimming routes if necessary to the exact desired distance. This approach has been implemented to power a web-based, route recommendation software that allows for the identification of preferences, selection of a route from a set of alternatives, and export of the route in the .gpx format for use with smartwatches.

11 - On-demand service capacity planning: a bi-level approach

Seokwoo Kim, Pohang University of Science and Technology (POSTECH), Pohang-si, Korea, Republic of, Seung Min Baik, Dong Gu Choi

This study addresses the capacity planning problem in on-demand service systems. Unlike conventional service systems, capacity uncertainty arises in the on-demand service systems due to high autonomy in servers. Recent on-demand platform studies have investigated strategic behaviors of servers based on game-theoretic approach. Expanding upon both traditional planning model and game-theoretic behavior model, this study proposes an integrated service capacity planning model by incorporating interactions among servers, as well as between servers and operators. The proposed model is formulated into a bi-level program with a single-leader multiple-follower structure. A pessimistic approach is chosen to mitigate the multiple equilibrium points. We show that the proposed bi-level model can be formulated as a two-stage stochastic copositive program.

12 - Three essays on multiple orders per job (MOJ) scheduling problems in two-stage permutation flowshop

Rohan Korde, Arizona State University, Tempe, AZ, United States, John Fowler, Lars Moench

My dissertation research (organized as three essays) focuses on machine scheduling problems encountered in a 300-mm wafer fabrication (fab) facility where semiconductor chips are manufactured in reentrant process flows on hundreds of tools. Front opening unified pods (known as FOUPs or jobs) transport 300-mm silicon wafers between the different tools using an automated overhead hoist transportation system (OHT). The number of FOUPs are typically far fewer than the number of customer orders received. In some cases, customer orders can be combined into the same FOUPs depending on certain factors. The FOUPs are then scheduled to be processed on tools to minimize a given performance measure. These scheduling problems are known as "multiple orders per job" (MOJ) problems and are the focus of my dissertation research. The first essay covers single objective MOJ scheduling problems to minimize makespan, total completion time, and total weighted completion time for orders with equal and unequal release times in a two-stage permutation flowshop. The second essay covers single objective MOJ scheduling problems to minimize maximum lateness, number of tardy orders, weighted number of tardy orders, total tardiness, and total weighted tardiness for orders with equal and unequal release times in a two-stage permutation flowshop. The third essay covers multiobjective MOJ scheduling problems to minimize pairs of uncorrelated performance measures for orders with equal and unequal release times in a two-stage permutation flowshop. In these essays we design full-factorial experiments to check optimality (small-sized instances) and solution quality (large-sized instances).

13 - A Stability Principle for Learning under Non-Stationarity

Chengpiao Huang, Columbia University, New York, NY, United States, Kaizheng Wang

We develop a versatile framework for statistical learning in non-stationary environments. In each time period, our approach applies a stability principle to select a look-back window that maximizes the utilization of historical data while keeping the cumulative bias within an acceptable range relative to the stochastic error. Our theory showcases the adaptability of this approach to unknown non-stationarity. The regret bound is minimax optimal up to logarithmic factors when the population losses are strongly convex, or Lipschitz only. At the heart of our analysis lie two novel components: a measure of similarity between functions and a segmentation technique for dividing the non-stationary data sequence into quasi-stationary pieces.

14 - Modeling TiA in Teams: Leveraging Acoustic and Lexical Features in Dynamic Trust Estimation

Morgan Klaeser, University of Wisconsin-Madison, Madison, WI, United States, Vianney Renata, John Lee

Traditionally, trust in automation has been characterized as a function of an individual's cumulative experience. Recent research highlights the role of team environments in shaping dynamic trust. Teams foster feedback loops and cultivate affective environments that influence cognitive processes underlying trust in automation.

This study assesses the suitability of a Bayesian Hierarchical Potts model in capturing trust dynamics within teams, using acoustic and lexical features to describe the team environment. The Potts model, from statistical mechanics, represents systems with interacting components and can capture phase transitions and clustering behavior. Inspired by prior work on the NASA Hera Study, we estimate trust dynamics in team contexts.

We analyze six weeks of naturalistic audio data from teams using an automated conversational agent during mission practice, examining experimental and social conditions linked to trust outcomes. Acoustic and lexical features describe the team environment, influencing the phase an individual occupies within the model. Likert scale ratings correspond to distinct phases, representing varying trust levels.

Our findings use a novel methodology for long-term trust characterization, quantifying the role of affective environments in cognitive trust formation. We show how social-emotional environments significantly impact trust in automation. Ignoring such a social environment risks latent variable bias in trust dynamics.

By quantifying the relationship between trust and team environment, our study offers insights for designing feedback systems to enhance team performance and ensure mission success. These insights extend to other environments where trust in automation is crucial, promoting smarter decision-making and a safer integration of automation in society.

15 - ESG Data Imputation and Greenwashing

Giulia Crippa, Princeton University, Princeton, NJ, United States

In recent years, there has been a notable surge of Environmental, Social, and Governance (ESG) investing. This paper provides a simple and comprehensive tool to tackle the issue of missing ESG data. Firstly, it allows to shed light on the failure of ESG ratings due to data sparsity. Exploiting machine learning techniques, we find that the most significant metrics are promises, targets and incentives, rather than realized variables. Then, data incompleteness is addressed, which affects about 50% of the overall dataset. Via a new methodology, imputation accuracy is improved with respect to traditional median-driven techniques. Lastly, exploiting the newly imputed data, a quantitative dimension of greenwashing is introduced. We show that when rating agencies do not efficiently impute missing metrics, ESG scores carry a quantitative bias that should be accounted by market players.

16 - Looking into the future: Job insecurity and future work selves as a moderator

Vivien Lim, National University of SIngapore, Singapore, Singapore, Thompson Teo

Studies across various fields have established the importance of positive possible selves in influencing a wide range of individual behaviors, from reducing in-school misbehaviors to increasing the frequency of engaging in exercise (e.g.,Oyserman et al., 2006). Due to the changing employment landscape, there is a growing scholarly interest in the study of possible selves specific to the work domain (Strauss, Griffin, & Parker, 2012). Termed future work selves, these possible selves embody individuals' hopes and aspirations with respect to their future careers.

The current study examined the motivational value of the future work self in the context of a work-related stressor, specifically job insecurity. We examined the moderating effect of salient future work selves and job insecurity on work-related outcomes among mid-career individuals. Data from this study were collected from 182 police officers and their immediate supervisors. Findings largely supported the importance of salient future work selves in buffering individuals from the negative effects of adversities, through increasing their motivation and confidence to overcome challenging situations.

In investigating the utility of the future work selves as a psychological resource in the job insecurity context, this study has important practical implications for employees as it provides evidence of future work selves as a valuable resource which individuals can draw upon when their job security is threatened. As individuals are increasingly required to manage their own careers, the future work self is an exciting construct that can help employees regulate their behaviors through stressful times.

17 - Effects of Heterogenous Decisionmakers in Disrupted Supply Chains

Souri Sasanfar, Northeastern University, Boston, MA, United States

Title:

"Effects of Heterogenous Decisionmakers in Disrupted Supply Chains"

Disruptions to the supply chain present serious problems, especially in the pharmaceutical industry where reliable and effective delivery is essential. This study examines the effects of decision-maker heterogeneity on supply chain performance during different types of disruptions, with a particular emphasis on three important profiles: followers, hoarders, and reactors. To study the dynamic interactions between these decision-makers and two important information sharing formats, Automated System Information (ASI) and Automated Demand Information (ADI), we created an agent-based simulation model.

The simulation looks at how different decision-maker types react to supply chain interruptions and how their actions interact with ADI and ASI to impact performance. By lowering supply chain expenses, panic purchases, and phantom ordering, the study seeks to determine which information sharing practices could enhance performance and offer suggestions for the best kinds of ADI and ASI for pharmaceutical supply chains.

The goal of the research is to provide a comprehensive understanding of the actions and tactics used by various decision-makers, along with recommendations for improving collaboration and decreasing inefficiencies. By doing so, stability in this important industry will be ensured and pharmaceutical supply chains will become more robust and effective.

18 - Anomaly Detection in Connected and Autonomous Vehicle Trajectories Using LSTM Autoencoder and Gaussian Mixture Model

Boyu Wang, Beijing University of Civil Engineering and Architecture, Beijing, China, People's Republic of, Wan Li, Zulqarnain Khattak

Connected and Autonomous Vehicles (CAVs) technology has the potential to transform the transportation system. Although these new technologies have many advantages, the implementation raises significant concerns regarding safety, security, and privacy. Anomalies in sensor data caused by errors or cyberattacks can cause severe accidents. To address the issue, this study proposed an innovative anomaly detection algorithm, namely the LSTM Autoencoder with Gaussian Mixture Model (LAGMM). This model supports anomalous CAV trajectory detection in the real-time leveraging communication capabilities of CAV sensors. The LSTM Autoencoder is applied to generate low-rank representations and reconstruct errors for each input data point, while the Gaussian Mixture Model (GMM) is employed for its strength in density estimation. The proposed model was jointly optimized for the LSTM Autoencoder and GMM simultaneously. The study utilizes realistic CAV data from a platooning experiment conducted for Cooperative Automated Research Mobility Applications (CAR-MAs). The experiment findings indicate that the proposed LAGMM approach enhances detection accuracy by 3% and precision by 6.4% compared to the existing state-of-the-art methods, suggesting a significant improvement in the field.

19 - A unified uncertainty-informed approach for risk management of deep learning models in the open world

Long Xue, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Sai Ho Chung, Lechang Yang, Xiaolin Wang, Xiaoge Zhang

Equipping deep learning models with a principled uncertainty quantification (UQ) has emerged as a cornerstone for safe and reliable adoptions of deep learning models in the open world. To handle uncertainty arising from two prevailing sources - distribution shift and out-of-distribution (OOD) - in the open-world setting, this paper develops a unified uncertainty-informed approach to assess and manage the risk of deep learning models. To this end, we propose to leverage a principled UQ approach – Spectral-normalized Neural Gaussian Process (SNGP) – to quantify the epistemic uncertainty associated with the model's predictions. Different from other UQ methods in the literature, the SNGP features two unique properties: (1) normalizing the weights of each hidden layer with spectral normalization to preserve the relative distance among samples during feature transformations; (2) replacing the traditional output layer in the neural network with Gaussian process for distance-aware uncertainty estimation. Based on SNGP, we divide the uncertainty into three levels by setting appropriate uncertainty thresholds to help decision-makers make risk-informed decisions. A case study on diabetes prediction is used to demonstrate how the proposed method facilitates risk assessment and management of deep learning models in the open environment. Computational results reveal that the proposed method achieves a prediction performance comparable to Monte Carlo dropout and deep ensemble methods. More importantly, the proposed method produces much more consistent and principled uncertainty estimations under no distribution shift, distribution shift, and OOD than the other methods in the literature.

20 - Color-Coding the Hydrogen Renaissance through a Prism of Hydrogen Production Technologies: Technology selection scenario analysis through bottom-up model construction

ChaeHyeon Kim, Seoul National University, Seoul, Korea, Republic of, Yoonmo Koo

As the era of hydrogen arrives, an innovative approach is needed to systematically prepare and promote the hydrogen Renaissance. This paper highlights the importance of hydrogen production technology during this transition period and analyzes its complexity by classifying various existing hydrogen production technologies by color. In this study, a bottom-up technology selection simulator named "Prism" is developed to analyze the selection weight of each technology under various scenarios, encompassing major hydrogen production technologies such as gray, blue, green, and yellow.

The innovation of this paper is to derive spectral results based on input data for each scenario in color form through the "prism" simulator. This method allows visual identification of the most suitable proportion of hydrogen production technologies under given conditions and scenarios through the colors representing each hydrogen production technology. This provides intuitive and clear guidelines for policymakers and industry in establishing sustainable hydrogen production strategies.

As a result, this study presents an innovative method of color-expressing the optimal technology combination and strategies for sustainable hydrogen production by comprehensively comparing and analyzing various aspects of hydrogen production technology. This study is expected to effectively visualize the economic, environmental, and technical characteristics of each technology and make an important contribution to establishing a policy direction for the sustainable development of hydrogen production technology through the application of a color code system that encompasses existing major hydrogen production technologies.

21 - How the Domain-Adaptation Language Model Works in Government Domain Section

jiwon kim, Kyung Hee University, Seoul, Korea, Democratic People's Republic of, Sunghyun Yoo, Sangjun Lee, yurianna Kim, Jae Hong Park

According to an industry survey, 55 percent of people lost their opportunities to apply for the programs because they don't know how to apply and what to need. Specially, people need to search and manually check whether such programs are fitted with them. In addition, many users have a difficulty to find information about what types of beneficiaries in the policy documents. Recently, a language model is regarded as a great tool and has shown a great performance in various tasks including document summarization (e.g., question-answering, summarizing, and understanding passages) to help people understand easily. So, we argue that using language model can give important information and opportunity to people who don't know how to apply a program. We believe that the language model with domain adaptation will make people to search clear information easily. So, this study proposes the language model to deeply understand the policy documents so that out model can help people search them easily. To do so, this study use government policy documents and domain NER data to train tokenizer and language model. In particular, this study develops the Auto Encoder language model 110,589 public policy data set. Our results, shows that domain adaptation significantly improves model performance. So, we believe that this study can contribute to the related literature in which this study conducts the tasks in government policy domain. Also, we believe that many people can easily search and find government subsidy documents to get more benefits.

22 - Data Utilization for Operational Efficiency and Customer Satisfaction in Gas Appliance Repair Services

Yuichiro Tanaka, TOKYO GAS CO.,LTD, Tokyo, Japan

Tokyo Gas Group serves around 9 million customers in the metropolitan area with electricity, gas, and services. Following the liberalization of Japan's electricity and gas retail markets in 2016 and 2017, expanding services beyond energy retail has become increasingly important.

TGG handles approximately 300,000 repair jobs annually for gas equipment. The reliance on individual skills in repair operations caused variability in service levels, leading to issues such as revisits to customer homes due to missing repair parts. This inconsistency indicates room for improvement in both operational efficiency and customer satisfaction.

To address these challenges, TGG overhauled the entire repair service and implemented data-driven operational transformations. We will introduce three key initiatives:

1. **Predict Fault Parts Before Visits**
- By developing a model to predict repair parts before visits, we significantly improved prediction accuracy over the current model. This suggests a potential to exceed the KPI of One-Time Completion Rate by 8 to 26 percentage points.

2. **Optimize Store Inventory Management**
- Using machine learning and statistical methods, we optimized inventory management. This indicated the potential to improve the Repair Parts Availability Rate by 30 percentage points and reduce excess inventory.

3. **Monitor Customer Satisfaction**
- Customer satisfaction was enhanced by monitoring and analyzing post-repair surveys. This resulted in a 5-point improvement in overall service satisfaction within one year.

23 - Defining abnormal trends and identifying abnormal wafers using object detection methods

kuhyun lee, Sungkyunkwan University, suwon, Korea, Republic of, Dong-Hee Lee

In a semiconductor wafer fabrication process, it is critical to identify abnormal wafers and to analyze the causes to improve the process. Because a massive number of wafers are fabricated, a large number of wafers are often inspected at the same time. In this case study, we assume a quality control system which displays quality inspection results of several wafers in a single chart. The purpose of this study is to define abnormal trends observed from the charts and further identify abnormal wafers that contribute to the trends. For this purpose, we employed object detection methods such as YOLO. Once the abnormal wafers are identified, their process log data are analyzed by the quality inspection system, so-called commonality analysis, to identify faulty machines that may have resulted in the abnormal wafers. The case study proved that the proposed definition of the abnormal trend is valid and the abnormal wafers identification method works well.

24 - BBScore: A Brownian Bridge Based Metric for Assessing Text Coherence

Chen Jiang, University of Minnesota, Minneapolis, MN, United States, Zhecheng Sheng, Tianhao Zhang, Dongyeop Kang

Measuring the coherence of text is a vital aspect of evaluating the quality of written content. Recent advancements in neural coherence modeling have demonstrated their efficacy in capturing entity coreference and discourse relations, thereby enhancing coherence evaluation. However, many existing methods heavily depend on static embeddings or focus narrowly on nearby context, constraining their capacity to measure the overarching coherence of long texts.
In this paper, we posit that coherent texts inherently manifest a sequential and cohesive interplay among sentences, effectively conveying the central theme, purpose, or standpoint. To explore this abstract relationship, we introduce the "BBScore," a novel reference-free metric grounded in Brownian bridge theory for assessing text coherence. Our findings showcase that when synergized with a simple additional classification component, this metric attains a performance level comparable to state-of-the-art techniques on standard artificial discrimination tasks.
We also establish in downstream tasks that this metric effectively differentiates between human-written documents and text generated by large language models under a specific domain. Furthermore, we illustrate the efficacy of this approach in detecting written styles attributed to diverse large language models, underscoring its potential for generalizability. In summary, we present a novel Brownian bridge coherence metric capable of measuring both local and global text coherence, while circumventing the need for end-to-end model training. This flexibility allows for its application in various downstream tasks.

25 - RealTwin: An Automated Scenario Generation Tool for Traffic Simulation

Guanhao Xu, Oak Ridge National Laboratory, Oak Ridge, TN, United States, Abhilasha Saroj, Ross Wang, Yunli Shao

Traffic simulation has been a valuable approach for urban planners, traffic engineers, and researchers to study traffic. Particularly, microscopic traffic simulation, which simulates individual vehicles' movements within a transportation network, has demonstrated its importance in analyzing and managing transportation systems. However, it has always been a challenge to integrate data from various sources, generate traffic scenarios, and import it into traffic simulators to conduct microscopic simulations. This paper provides solutions to overcoming this challenge. Specifically, RealTwin, a comprehensive tool for automated scenario generation for microscopic traffic simulation is presented in this paper. By providing a streamlined scenario generation workflow, RealTwin effectively bridges gaps between traffic data from various sources and traffic simulators, making microscopic traffic simulation more accessible for researchers and engineers across various levels of expertise. By using RealTwin to generate a real-world traffic scenario in AIMSUN and SUMO, the ability of RealTwin to construct comparable simulations in different simulators that lead to consistent simulation results is demonstrated. Furthermore, we introduce and illustrate RealTwin's extended capability for technology (e.g. autonomous vehicle) scenario generation. This extended feature can contribute to more comprehensive microscopic simulations, facilitating the analysis of the potential impacts of various technological innovations on mobility, energy efficiency, and safety.

26 - Optimizing Active Distribution Grids with Weather- and Decision-Dependent Reliability

Gejia Zhang, Rutgers University, Piscataway, NJ, United States, Robert Mieth

The reliability of power systems is critically influenced by weather conditions, which directly affect the operational integrity of physical infrastructure and electricity consumption patterns. Emerging smart grid technology enables system operators to react to temperature fluctuations, but tools to quantify and internalize failure risk on a component-level in operative decisions are still lacking. In this work we develop a distribution system model to optimize the control set-points of distributed generation units, battery energy storage systems, and demand response resources, that explicitly considers weather- and decision-dependent system reliability. Through decision-dependent reliability, our model captures how operational decisions and ambient temperature impact the likelihood of component failures, allowing operators to directly balance cost efficiency and reliability. To address the non-linear and non-convex optimization challenges of the model, we develop an iterative optimization strategy using linearization techniques around initial reliability estimations derived from a baseline scenario. We demonstrate our model and methodology using numerical case studies on IEEE test systems.

27 - Integrated supplier performance risk tool

chaoye pan, Ford Global Data Insight & Analytics, Northville, MI, United States, Oleg Gusikhin

For automotive industries, many potential threats are associated with working with suppliers, like quality issues, delivery disturbance, financial instability, and more. By identifying and assessing these risks upfront, implementation of an internal & external supplier performance risk management process will ensure supplier performance is embedded in end-to-end supply chain development & fulfillment activities. ISPR brings together QDC + Concentration & Compliance metrics in a single view to develop risk profiles for champions to understand the key risk and performance indicators associated, predictive analytics based on available internal & external supplier information and centralized location to prioritize case management, tracking and follow-up.

28 - SANO: a Sterilization Resource Allocation Tool via Network Optimization

Yuming Yang, Johnson & Johnson, Raritan, NJ, United States, Hanzhang Liu

In MedTech industry, a substantial portion of products undergo sterilization using Ethylene Oxide (EO) which is subject to stringent environmental regulations. Managing sterilization site capacity, optimizing utilization, and allocating products in the hierarchical supply chain network becomes critical to Multinational Corporations, given their global presence, production scale, and operation complexity. However, the traditional method of manual planning to configure network flow is suboptimal and uneven, resulting in extremely high utilization at certain sites while others are vacant. It is not only impractical for the long term, but also challenging to implement among teams when various business segments share capacity, leading to low efficiency and a lack of holistic perspective.

To address these supply chain challenges, we propose SANO: a Sterilization Resource Allocation Tool via Network Optimization. Specifically, we design a customized network optimization model in Python that accounts for real-life operational constraints and objectives. Furthermore, we deploy the model in a cloud-based environment and develop a user-friendly web interface with interactive dashboards, enabling scenario exploration and analysis to users without optimization background. Evaluation of real-world scenarios demonstrates a significant reduction of 50% in planner workload and complete elimination of capacity shortage, alleviating network bottleneck. Providing global optimal results, this solution minimizes total network costs, EO usage, and wastage. It also supports informed business decisions and aids in establishing a competitive and proactive supply chain network.

29 - Courier scheduling in last-mile delivery considering show-up rate dynamics and job proficiency

Yuin Moon, Pohang University of Science and Technology (POSTECH), Pohang, Korea, Republic of, Seokwoo Kim, Dong Gu Choi

In this study, we consider a courier scheduling problem for last-mile food delivery systems, considering characteristics of couriers; 1) uncertainty in the actual participation of scheduled couriers, 2) differences in job proficiency between couriers, and 3) uncertainty in customer demand. We formulate a two-stage mixed-integer stochastic programming model for our problem. The first stage presents the schedule of couriers during the planning phase, and the second stage comprises the adjustment decision during the operational phase. We develop a solution approach based on the progressive hedging algorithm. Furthermore, the effectiveness of the proposed model is validated using simulated environments based on real-world data.

30 - Development of an Integrated Mode and Route Choice Model for Shared Mobility in Minnesota

Mobina Nankali, University of Minnesota, Minneapolis, MN, United States, Micheal Levin

Shared mobility changes traffic congestion, infrastructure use, accessibility, and emissions. Predicting these changes is necessary for urban planners and policymakers to improve traffic management systems and strategize incorporating shared mobility into the existing infrastructure. By integrating a logit choice model with a user equilibrium (UE) model, the study addresses the growing complexities introduced by shared transportation options such as transportation network companies, bike sharing, and car sharing. The proposed model considers the fleet size limit as a constraint for certain modes of transportation. A function is introduced as a penalty added to the objective function to address this constraint. This approach separates the problem into two optimization problems, with the first one being a user equilibrium that is solved by a Bush-based algorithm. Findings from the model demonstrate its capability to accurately forecast changes in travel behavior and traffic patterns, highlighting the potential reductions in vehicle travel and greenhouse gas emissions through the adoption of bike sharing and electric car sharing.

31 - A link transmission model of eco-approach autonomous vehicles in mixed platoons to predict the impacts on energy consumption in city networks

Maziar Zamanpour, University of Minnesota, Twin cities, MN, United States, Michael Levin

Automated vehicles (AVs) have become highly valued due to the numerous potential benefits in transportation systems. AVs have unveiled promising opportunities to enhance safety, mobility, and efficiency. Utilizing technology and connectivity, speed control strategies have led to major energy efficiency by avoiding abrupt deceleration-acceleration cycles which we refer to as eco-approach AVs. Although speed control presents well-established energy benefits for AVs in limited scenarios, applying eco-approach speed control for AVs also affects any following human-driven vehicles (HVs) that are forced to drive at similar speeds.

This study is the first attempt to predict the network-level energy benefits of eco-approach AVs, including fuel benefits to HVs. A modified link transmission model (LTM) obtains traffic flow for each link in the network influenced by eco-approach AVs. HVs are modeled as either passing slower-moving eco-approach AVs or following their speed (and achieving speed benefits). An approximated but computationally faster speed control strategy is presented to mimic the eco-approach control (which typically involves solving model predictive control). A parametric longitudinal load model that incorporates the principles of vehicle dynamics estimates energy consumption which is reported for various AV penetration rates and also HVs. Moreover, since connected intersections play a crucial role in transmitting information for eco-approach AVs, we study the route choice of AVs for a limited number of equipped intersections (due to budget constraints). In addition to AVs, route choice and mobility for HVs are also compared for a test network.

32 - How to Compete with AI-Generated Contents?

Seyed Esmaeili, University of Chicago, Chicago, IL, United States

Generative AI will have significant impact on content creation platforms. In this paper, we study the dynamic competition between an AI and a human contributor. Unlike the human, the AI's content only improves by having access to more samples, however the AI has the advantage of generating content at a low cost. Given this model we study how the human contributor can maximize his utility when competing against the AI for content generation over a collection of topics. Specifically, we consider a long horizon setting and show a novel polynomial time algorithm to maximize the human creator's utility. In addition, we consider a setting where the samples have to be ``fresh" for high quality AI generation which is the case in many topics such as news and weather. Interestingly, we show that the problem of maximizing the human creator's utility in this setting is NP-hard. For this setting, we introduce a simple polynomial time algorithm that cycles between myopically optimizing over a short time window and pausing and show that it essentially has an approximation ratio of 0.5.

33 - Co-purchaser Choice in Threshold Group-Buying: Empirical Study with Panel Survival Analysis and Logit Model

JOOHWAN KIM, Kyung Hee University, Seoul, Korea, Republic of, yongjoon kwon, JAEHONG PARK

A group-buying is a method of purchasing goods together with a discounted price at platforms. In the group-buying platform, users can choose two roles – initiator or co-purchaser. The initiator starts the group-buying deal to buy goods together in the platform while the co-purchaser participates in the deal which the initiator already created. Since the deal can only be completed when the co-purchaser joins the deal, the participation of co-purchasers is essential to complete a deal on group-buying platforms. So, the success of the deal depends on the co-purchasers' choice. However, previous studies focus on the motivation of why people join the group-buying deals regardless of their roles - initiator or co-purchaser. Specially, few studies have examined the role of co-purchasers even though the co-purchasers are more important to complete the deal. So, this study investigates what factors make co-purchasers to join the deal with panel survival models. Specially, we examine the impact of the left time of the deal on co-purchasers' choice, which lead to the success of group-buying deals. Using real-time log data sets from one of the largest group-buying platforms in South Korea, we empirically show that the left time of the deal has a negative impact on its success of group-buying deals. Although the left time of the group-buying deals do not provide any utilities to the co-purchaser, we find that co-purchasers often join the deals with less time left. We suggest that co-purchasers might be motivated to purchase earlier due to their sense of urgency.

34 - Quality-based scheduling in a multistage manufacturing process using Monte Carlo tree search method

Sugyeong Lee, Department of Semiconductor and Display Engineering, Sungkyunkwan University, Suwon-si, Korea, Republic of, Dong-Hee Lee

In multistage manufacturing processes, quality can vary significantly depending on the chosen path, due to the difficulty of ensuring uniform product quality from each piece of equipment. We introduce the use of Monte Carlo tree search (MCTS) method to consider quality in addition to productivity in the path selection process. Specifically, we propose a method to incorporate both short-term and long-term benefits as rewards in the MCTS model by utilizing main effects and interaction effects from traditional experimental designs, to produce paths that maximize yield. This approach is expected to allow optimal scheduling while ensuring quality even in large-scale multistage process environments where conventional methods may struggle to estimate quality outcomes.

35 - <div max-w-[90%]"="">Optimal Battery Utilization and Swapping Service Level through Balanced Battery Inventory and Replenishment for an Electric Vehicle Battery Swapping Station Network

Reza Alizadeh, Magna International, Boston, MA, United States

Establishing an ample charging infrastructure is pivotal for the successful deployment of electric vehicles (EVs), mitigating concerns associated with restricted range and extended charging durations. Battery exchange stations arise as a promising remedy to alleviate these waiting periods. These stations act as intermediaries between the power grid and consumers, presenting a rapid alternative to recharging batteries. Ensuring their efficacy and profitability necessitates a robust business and operational model. This endeavor introduces an optimization framework for the operational structure of battery swapping stations, emphasizing balanced battery inventory management. The framework offers features like flexible battery restocking from depots and equitable battery inventory distribution among charging stations. The battery float percentage and battery utilization rate are augmented, eliminating unmet demand and minimizing the overall shipment cost. The findings demonstrate the feasibility of the proposed model as a business case, as well as its efficacy in delivering the necessary service.

36 - Patched Vulnerability Disclosure and Hacker Participation in Bug Bounty Programs

Ali Ahmed, Louisiana State University, Baton Rouge, LA, United States

In cybersecurity, bug bounty programs (BBP) have emerged as an effective method of finding hidden vulnerabilities in a system. In this method, ethical hackers discover and report hidden vulnerabilities to the organizations. In return, these hackers receive rewards for their valid discoveries. The emergence of bug bounty programs has enabled the practice of Patched Vulnerability Disclosure (PVD), where firms disclose vulnerability information after they have been patched. Currently, there is little understanding of how PVD affects ethical hackers' participation in BBPs. We collected a dataset from a leading bug bounty platform and conducted empirical analyses to analyze the effect of PVD on hacker participation. Leveraging multiple specifications, including an instrumental variable approach, difference-in-differences, and data collected from third-party websites, our analysis reveals that as firms disclose more patched vulnerabilities, hacker participation in their programs decreases. To further explore the mechanism, we found that disclosing valid vulnerability reports discourages hacker participation, whereas disclosing invalid reports leads to more hacker participation. We found that PVD influences uncertainty in the bug bounty process, which can simultaneously impact the risk and ambiguity in hacking. Valid disclosures can escalate hackers' risk and discourage participation, whereas disclosing invalid submissions can reduce the ambiguity in the bug bounty process and increase the hacker's participation in BBP. We also found that reduced ambiguity is most salient for hackers with experience working with other firms on the platform. Overall, our study makes theoretical contributions to the literature on vulnerability disclosure, ethical hacking, and the management of bug bounty programs.

37 - Methods, Analysis, and Insights from a State-Of-The-Art Large Glucose Model

Abhimanyu Kumbara, Welldoc, Inc., Columbia, MD, United States, Junjie Luo, Anand Iyer, Mansur Shomali, Guodong Gao

Proactive self-management of diabetes requires accurate glucose value prediction and automated, in-the-moment AI-driven coaching based on those predictions. We created a large glucose model (LGM) trained exclusively on millions of glucose values from continuous glucose monitoring devices (CGM). This model can achieve state-of-the-art (SOTA) results in predicting glucose trajectories in type-1 and type-2

diabetes (T1D, T2D) populations at 30min, 60min, and 2-hour time intervals. In the current analysis, we developed a new SOTA LGM that used both glucose values and time series inputs to predict glucose trajectories at the time intervals above. We also analyzed the results of the LGM across different population subgroups.

When comparing the average root mean square error (RMSE) in T1D and T2D populations associated with the hour of the predicted glucose values, the LGM performed better when predicting glucose trajectories between 12am – 5am, with RMSE increasing during breakfast (6am – 10am), lunch (12pm – 3pm), and dinner (5pm – 9pm). For the T1D population, the RMSE of the model was inversely correlated with age group, and for T2D population, the RMSE of the model was positively correlated with age group. When looking at total engagement levels across medication, education, diet, activity, and labs (MEDAL) categories, engagements between 20-30 entries per month minimized the RMSE for both populations. Analysis and insights from LGM enable us to understand subgroup level differences in managing diabetes and can be used to develop smarter personalized coaching applications that can ultimately be used to improve health outcomes.

38 - Distributed Learning Dynamics for Coalitional Games

Aya Hamed, University of Illinois Urbana-Champaign, Champaign, IL, United States, Jeff Shamma

In the framework of transferable utility coalitional games, a scoring (characteristic) function determines the value of any subset/coalition of agents. Agents decide on both which coalitions to form and the allocations of the values of the formed coalitions among their members. An important concept in coalitional games is that of a core solution, which is a partitioning of agents into coalitions and an associated allocation to each agent under which no group of agents can get a higher allocation by forming an alternative coalition. We present distributed learning dynamics for coalitional games that converge to a core solution whenever one exists. In these dynamics, an agent maintains a state consisting of (i) an aspiration level for its allocation and (ii) the coalition, if any, to which it belongs. In each stage, a randomly activated agent proposes to form a new coalition and changes its aspiration based on the success or failure of its proposal. The coalition membership structure is changed, accordingly, whenever the proposal succeeds. Required communications are that: (i) agents in the proposed new coalition need to reveal their current aspirations to the proposing agent, and (ii) agents are informed if they are joining the proposed coalition or if their existing coalition is broken. The proposing agent computes the feasibility of forming the coalition. We show that the dynamics hit an absorbing state whenever a core solution is reached. We further illustrate the distributed learning dynamics on a multi-agent task allocation setting.

39 - Transparent differential diagnosis using ensemble of simple structures models

Gaurav Arwade, Iowa State University, AMES, IA, United States, Sigurdur Olafsson

Heart failure remains a prevalent global health concern, with 6.5 million diagnosed cases in the United States alone and an annual increase of 960,000 new cases. The complexity of this clinical syndrome, often intertwined with various comorbidities, presents challenges in accurate diagnosis. Factors such as cardiac dysfunction stemming from conditions like ischemic heart disease, atrial fibrillation, valvular heart disease, and uncontrolled systolic blood pressure can contribute to undiagnosed heart failure. Conversely, symptom overlap with related conditions such as obesity and respiratory diseases can lead to both over and misdiagnosis, particularly in primary care settings where resources for comprehensive evaluation are limited. Breathlessness is one of the common symptoms between heart failure and its differential diagnosis; thus, primary patients presenting this symptom can be misdiagnosed. Predictive models may help improve diagnostic accuracy, but in this domain transparency and explainability of the models is also critical. To address these diagnostic challenges, we propose an ensemble approach utilizing simple models constructed upon underlying data structures. These models leverage subsets of patients exhibiting similar characteristics, where redundant complex feature interactions are minimized. While deep learning and complex algorithms like XGBoost have demonstrated impressive diagnostic performance, their lack of explainability has led to skepticism among physicians regarding recommendations from 'black box' algorithms. Our approach offers improved explainability without compromising accuracy, making it more accessible and trustworthy for clinicians. By harnessing the inherent simplicity within the data, our ensemble of simple models suggests a pragmatic solution for accurate heart failure prediction in primary care settings.

40 - Minimax Rate Optimal Algorithms for High-Dimensional Stochastic Linear Bandits

Jingyu Liu, Queen's University, Kingston, ON, Canada, Yanglei Song

We consider the stochastic linear bandit problem with multi-arms over T rounds, where the dimension d of the covariate vectors is potentially larger than T, however each arm-specific parameter vector has at most s_0 non-zero components. First, we study the single-arm case, focusing on the cumulative estimation error of the parameter vector. We show that Lasso has a sub-optimal dependence on d,T in terms of worst-case error in the online estimation setup. Further, we establish the minimax optimality of the OLS post Lasso, which first estimates the support of the parameter vector by thresholding the Lasso estimator and then runs least squares on the selected support. Second, we consider the bandit setup, using the OLS post Lasso as the main estimation method. We propose a three-stage algorithm for arm selection and derive an upper bound of order $s_0 (\log s_0) (\log d + \log T)$ on its cumulative regret. Finally, we establish a matching lower bound up to a $\log s_0$ multiplicative term, which shows the near minimax optimality of the proposed method. We note that our analysis avoids the restrictive "beta-min" condition.

41 - Enhanced Detection of Unknown Defect Patterns on Wafer Bin Maps Based on Open-Set Recognition Approach

jinsu shin, Department of Semiconductor and Display Engineering, Sungkyunkwan University, Suwon-si, Korea, Republic of, Dong-Hee Lee

Detecting and classifying defect patterns on wafers in semiconductor manufacturing processes is crucial for wafer quality management and prompt analysis of defect causes. However, due to the continuous technological innovation and advancement in semiconductor industry processes, the likelihood of unknown defect patterns emerging is increasing. These unknown defect patterns present a challenge that traditional classification and clustering methods, relying solely on existing training data, struggle to effectively address. To overcome this challenge, this study proposes a novel methodology based on Open-Set Recognition (OSR) using EEOC-SVM (Entropy Estimation One-Class SVM) to accurately detect unknown defect patterns. This methodology introduces unique preprocessing steps, including C-Mean filtering and Radon transformation, to remove noise and efficiently extract features from wafer bin maps (WBM). Through this approach, the study demonstrates the successful detection of unknown defect patterns in test data using only known defect pattern data. These results validate the proposed method as a robust tool for effectively detecting and addressing unexpected defect patterns, thereby expecting to significantly impact quality management and maintenance in future semiconductor manufacturing processes.

42 - Optimal Sensor Placement Using Denoising Autoencoder and Discrete Optimization for Heat Transfer Digital Twins

Chaeyun Yeo, Kyung Hee University, Department of Artificial Intelligence, Giheung-gu, Korea, Republic of, Jin-Gyun Kim, Younghoon Kim

This study introduces a methodology for Optimal Sensor Placement (OSP) in constructing digital twins of heat transfer systems. By combining Denoising Autoencoder (DAE) with discrete optimization techniques, our approach aims to accurately reconstruct real-time heat transfer simulation results using a limited number of sensors. The OSP optimization employs a projection function driven by discrete optimization. This identifies the minimal sensor array necessary for precise temperature monitoring, thereby reducing installation and operational costs in industrial systems. Additionally, the DAE effectively processes noisy data, enhancing the robustness of OSP decisions. Comparative evaluations against previous methods demonstrate the superior efficiency of the DAE in processing thermodynamic data. Our proposed approach is expected to significantly enhance the performance of temperature monitoring and control system designs within heat transfer digital twins.

43 - Electrons to electrons? Matching objectives with stakeholder-driven metrics for equitable power systems planning

Ogechi Vivian Nwadiaru, University of Massachusetts, Amherst, Amherst, MA, United States, Paola Furlanetto, Krista Harper, Erin Baker

With the energy transition at the forefront of policy objectives, it is important to ensure that the system planning models capture community needs and are guided by equitable metrics. We develop a game for eliciting these metrics from various stakeholders. The game, which is inspired by apples to apples, calls on stakeholders to match a set of energy equity metrics to a predefined list of objectives, and a judge selects the most favorable metric. This work extends the hierarchical framework of value-focused thinking by actively involving stakeholders in the decision-making process.

44 - Analysis of Power System Impacts Due to Power-to-Heat Energy Conversion

Ye Ha Yang, Korea university, Seoul, Korea, Republic of

To achieve the 2050 carbon neutrality target and accelerate the deployment of renewable energy, power system issues such as curtailment, and the rapid electrification of heat demand due to distortion in the relative prices of energy causes excessive demand, exacerbating power system problems. To address these issues, the conversion between electric and heat energy can induce fluctuations in electricity demand, acting like a flexible resource. Currently, a significant portion of South Korea's electricity demand includes heat energy demand; therefore, separation of electric and heat energy is necessary for efficient energy consumption. However, study in the context of Korea is still lacking. Thus, it is necessary to analyze the effects of power-to-heat conversion for future policy implementation. For this purpose, a power system model capable of predicting generation, surplus electricity, system operating costs, and carbon emissions according to future scenarios in Korea was developed. The effects of the power-to-heat scenarios on the economic, environmental, and system stability aspects of the power system were calculated. Applying a 10% power-to-heat conversion by 2035 resulted in a reduction of fuel costs by about 5% economically, and a decrease in output limitation rates in terms of system stability. These results demonstrate that power-to-heat conversion can reduce base load and be used as a flexible resource in an increasingly unstable power system due to the continuous increase in renewable energy and the electrification in various industries.

45 - You Are What You Like: the Relations Between Daily Preferences and Higher Order Risk Attitudes

Peng Du, Southwestern University Of Finance And Economics, Chengdu, China, People's Republic of, Lin Zhao, Hua Chen

Higher order risk attitudes, including risk aversion, prudence, and temperance, play a fundamental role in economic decision-making, significantly influencing daily consumption/saving behaviors, portfolio allocation, and retirement planning. Their importance has stimulated a growing interest in gauging individuals' risk preferences using easily observable and reliable data. While extant literature has suggested that some easily identifiable basic daily preferences, such as specific diets and exercise preferences, can reflect an individual’s risk attitudes, theses relations have not been thoroughly investigated, with existing evidence predominantly restricted to the second-order risk preference. We, therefore, conduct a systematical survey among University students, encompassing thirty-six basic daily preferences across six common categories (food, clothing, exercise, reading, music, and video), along with three risky lottery-choice experiments to separately assess subjects' intensity of risk aversion, prudence and temperance. Our results indicate that, even after accounting for individual characteristics, certain basic daily preferences exhibit robust correlations with higher order risk attitudes. We also find that cognitive ability and personal traits serve as important mediating factors. Our study are the first to systematically investigate the relations between basic daily preferences and higher order risk attitudes. Additionally, our research provides a valuable demonstration of the potential for identifying risk attitudes through the analysis of individuals’ behavioral data in the era of big data.

46 - Deep Galerkin Method for Mean Field Control Problems

Jingruo Sun, Stanford University, Stanford, CA, United States

We consider an optimal control problem where the average welfare of weakly interacting agents is of interest. We examine the mean-field control problem as the fluid approximation of the N-agent control problem with the setup of finite-state space, continuous-time, and finite horizon. The value function of the mean-field control problem is characterized as the unique viscosity solution of a Hamilton-Jacobi-Bellman (HJB) equation in the simplex. Adapted from the classical Galerkin method, we construct a Deep Galerkin Method (DGM) to solve the HJB equation. Instead of seeking the reduced form of linear combinations, DGM approximates the solution using a deep neural network which is trained to satisfy the differential operator and boundary conditions of the high-dimensional nonlinear PDE. The accuracy of DGM results is validated through various numerical experiments. We also establish theoretical support by proving the convergence result for DGM. Specifically, we show that the loss function can be made arbitrarily small given that the value function possesses sufficient regularity. Based on that, we further demonstrate that if the loss function converges to zero, the corresponding neural network approximators must converge uniformly to the true value function on the simplex. Therefore, we propose a comprehensive framework for the high-dimensional mean-field control problem, leading to a reliable and efficient methodology for decision-making.

47 - Data Value in Distribution System Operations

Mehrnoush Ghazanfariharandi, Rutgers University, Piscataway, NJ, United States, Robert Mieth

The rise of advanced data technologies in electric power distribution systems enables operators to optimize operations but raises concerns about data security and consumer privacy. Resulting protection mechanisms that alter or obfuscate datasets may invalidate the efficacy of data-driven decision-support tools and impact the value of these datasets. This paper derives tools for distribution system operators to enrich data-driven operative decisions with information on data quality and, simultaneously, assess data usefulness in the context of this decision. To this end, we derive an AC optimal power flow model for radial distribution systems with data-informed stochastic parameters that internalize a data quality metric. We derive a tractable reformulation and discuss the marginal sensitivity of the optimal solution as a proxy for data value. Our model can capture clustered data provision, e.g., from resource aggregators, and internalize individual data quality information from each data provider. We use the IEEE 33-bus test system, examining scenarios with varying photovoltaic penetration, to demonstrate the application of our approach and discuss the relationship between data quality and its value.

48 - A (not so) Typical Day at Work for an Emergency Medical Services Crew Member

Nico Ornelas, Wichita State University, Wichita, KS, United States, Laila Cure

Emergency Medical Services (EMS) play a vital role in providing care to patients in need of urgent medical attention outside of the hospital. Most Operations Research/Management Science (ORMS) research on EMS systems focuses on optimizing operational decisions, such as station location, dispatching, and redeployment, in terms of some system-level performance measure, such as coverage or response time reliability. The impact of these decisions on individual crewmembers is rarely considered. Nevertheless, the tasks performed by these crews are extremely diverse and vary widely from shift to shift. This research uses data analysis techniques to describe the dynamics of EMS systems from the perspective of crewmembers. We start with a qualitative analysis of a single workday to use as a baseline. Then, we analyze one year's worth of emergency dispatch data to investigate deviations from this baseline and explore the amount of variation in the frequency and duration of the tasks performed. In addition, we use probability theory to analyze documentation expectations of ambulance crewmembers. The descriptive analysis provides a shared understanding of the work performed by paramedics and emergency medical technicians and the probabilistic analysis provides insight on how likely these crewmembers are to meet the articulated and unarticulated expectations of their profession. We conclude the analysis with future research directions for ORMS professionals as a way to consider the experience of crew members in the modeling and analysis of EMS operational decisions.

49 - End-to-End Asset Allocation Optimization with Convolutional Neural Network and Stock Chart Images

Inyeol Choi, Kyung Hee University, Giheung-gu, Korea, Republic of, Younghoon Kim

Optimizing dynamic asset allocation poses significant challenges but offers promising avenues for portfolio management. While prior research has utilized Convolutional Neural Networks (CNNs) to analyze stock chart images and forecast price movements within portfolios, our study introduces a new extension. We propose that chart images can effectively predict not only individual stock performance but also optimal asset allocation ratios. Our methodology entails developing a CNN-based end-to-end portfolio model trained on diverse allocation strategies. Through out-of-sample testing, the model autonomously generates optimal allocation weights without relying on predefined optimization processes. Leveraging visual data, our approach aims to enrich dynamic asset allocation strategies. This study highlights the potential of CNN-based models to effectively enhance portfolio management practices.

50 - Hierarchical Forecasting for Amazon EC2 demand

Yiming Wang, Amazon, Seattle, WA, United States, Ebrahim Nasrabadi

Amazon's Stores, Digital, and Other (SDO) is the largest user of AWS EC2 capacity. This poster introduces our in-house Hierarchical Bayesian and Hierarchical Reconciliation forecasting methods to forecast SDO EC2 demand. The demand forecasts are published at bi-monthly cadence, covering a rolling 2-year horizon, and at granularity of business group (Stores, Digital, Other), availability zone (e.g., us-east-1a), and instance type (e.g., c5.12xlarge) level. The forecasts can be used for AWS capacity planning to drive operational and financial decisions on global AWS data center infrastructure and space plan. To address the cold-start challenges in forecasting new instance generation and availability zones's demand, we developed a novel Hierarchical Bayesian time series forecasting model. In back-testing, our model was proven to reduce the forecast error measured in Weighted Absolute Percentage Error (WAPE) to 18.6% from 30.0% using Meta's open-source Prophet package for new AWS EC2 instances launched in 2023.

51 - Determinants of Patterns of Outpatient Care in Patients with Chronic Kidney Disease

Somi Kim, Seoul National University, Seoul, Korea, Republic of, Song-Hee Kim, Hyung Woo Kim

Regular outpatient visits by patients with chronic kidney disease (CKD) are crucial to manage patients' prognosis. Although the Kidney Disease Improving Global Outcomes (KDIGO) guidelines provide appropriate testing frequencies based on the patient's stage of CKD, actual visit frequencies do not strictly follow these guidelines due to influencing factors of outpatient visit patterns. This study conducted a retrospective analysis comparing visit frequency of the outpatient visit patterns of patients with CKD to the guideline and identified factors impacting visit frequency.

52 - Neural Forecasting Layer: Straightforward connected interpretable neural forecaster for multivariate time-series.

Wonkeun Jo, Chungnam National University, Daejeon, Korea, Republic of, Dongil Kim

Within the machine learning community, diverse neural network-based methodologies deployed in multivariate time-series forecasting applications have demonstrated commendable efficacy. Despite advancements in model performance, comprehending the rationale behind the model's behavior remains an enigma. Our proposed model, the Neural Forecasting Layer (NFL), employs a straightforward amalgamation of neural networks. This uncomplicated integration ensures that each neural network contributes inputs and predictions independently, devoid of interference from other networks. Consequently, our model facilitates a transparent explication of forecast results. This paper introduces NFL along with its diverse extensions. Empirical findings underscore NFL's superior performance compared to nine benchmark models across 15 available open datasets. Notably, NFL not only surpasses its competitors but also provides elucidation for its predictions. In addition, rigorous experimentation involving diverse model structures bolsters the justification of NFL's unique configuration.

53 - Tree ensemble-based data denoising technique for sparse dynamic system identification

Junhyeok Choi, Kyung Hee University, Giheung-gu, Korea, Republic of, Younghoon Kim

SINDy (Sparse Identification of Nonlinear Dynamics) is a widely used method for identifying nonlinear dynamical equations from data. However, its sensitivity to data noise presents challenges, especially in noisy environments like manufacturing processes. This study proposes a novel approach that combines denoising filters with tree ensemble algorithms to enhance robustness against noise and facilitate sparse equation identification. Comparative evaluations against basic SINDy demonstrate the superior robustness of our method when applied to various dynamical equations with Gaussian noise. These findings suggest that our approach shows promise for obtaining more reliable results in noisy environments, with potential applications in system identification and control. By mitigating the impact of noise, this study contributes to advancing the applicability of data-driven engineering methods in real-world settings, particularly in improving manufacturing processes and other practical applications.

54 - End-to-end portfolio optimization model using transformer

Jiyeon Lee, Kyung Hee University, Giheung-gu, Korea, Republic of, Younghoon Kim

In quantitative finance, portfolio optimization traditionally involves separate prediction and optimization stages, potentially leading to sub-optimal asset allocation decisions due to inaccuracies in parameter estimation. Recent advancements have introduced end-to-end portfolio optimization models, consolidating these steps into a single feedforward neural network. Our study extends this approach, focusing on a model-based strategy to enhance explainability and performance. We propose a method leveraging a transformer model for constructing an end-to-end portfolio optimization framework. We evaluate the performance of these models in maximizing the Sharpe ratio using U.S. financial data. Comparative analysis demonstrates the efficacy of our expanded model in improving portfolio optimization outcomes. This research contributes to advancing portfolio optimization techniques in quantitative finance, offering insights into more effective decision-making processes.

55 - Combining Shortcut3-ResNet and Gated Attention BiLSTM for Wafer Defect Classification

GYU MIN KANG, Kyung Hee University, Giheung-gu, Korea, Republic of, Younghoon Kim

Classifying wafer map defect patterns is crucial in semiconductor manufacturing to increase production yield. While previous methods rely on image analysis models like convolutional neural networks, they face performance limitations. To address this, we propose a composite model that integrates image and sequence information. Wafer image data undergoes preprocessing by resizing and distributionally gridding to facilitate model processing. We employ a combination of Shortcut3-ResNetBlock and MaxPool2d to transform high-dimensional image data into low-dimensional features. These features are processed using Gated Attention BiLSTM, enabling precise defect discrimination by merging image and sequence data characteristics. Our method capitalizes on the strengths of both data types, enhancing defect identification accuracy. FocalLoss is used in the learning process to effectively manage class imbalance. Experimental results demonstrate the superiority of our approach in achieving higher classification accuracy compared to existing methods. This advancement holds promise for significantly improving defect detection efficiency in semiconductor wafer fabrication processes, ultimately enhancing production yield.

56 - Development of One to One Marketing using AI in an Energy utility company

Takahiro Sekoguchi, Tokyo Gas, Tokyo, Japan, Shoma Hino, Kazuki Isshiki, Toshinori Sasaya

It has become more important for Japanese energy retailers to improve customer experience in the electricity and gas liberalization market.

Our marketing activities have faced challenges in understanding customers' needs, offering the optimal solutions, and communicating from customers' perspective.

We developed functions that integrate and analyze various types of customer data, estimate effective information for marketing using machine learning, such as customers' values, product needs and life events, and recommend products that meet the needs of each individual customer at the appropriate time and through the appropriate channels with customized communication tailored to individuals.

We verified the effectiveness of these marketing functions through individual use case such as estimating the timing of relocation and appealing electricity and gas bundle contracts, achieving significant improvement in conversion compared to our previous marketing methods.

We will systematize these marketing modules to deploy in both B2C and B2B, progressing with business implementation.

57 - A data-driven optimization for customized shuttle bus on the metro network with well-developed branch lines

Jiguang Wang, The University of Hong Kong, Hongkong, China, People's Republic of, Changyue Xu

Metropolitan subways are facing a serious connectivity issue among branch lines in the process of continuous development and expansion, especially at stations in the sub-centers of branch lines or at transfer stations, which leads to long travel distances frequent transfers. These problems not only reduce passenger travel efficiency and satisfaction but also burden rail transit operation agencies. Additional train shifts are often used on-site to alleviate the problem, but this method has limitations due to the inherent number of trains and the maximum departure frequency, and it may also worsen the supply-demand mismatch of other off-peak stations, resulting in high-cost waste. To tackle this problem, this study concentrates on metropolitan subway stations with great passenger demand between branch lines by utilizing big data mining and introducing a new customized subway-shuttle bus mode. The proposed mode aims to delivery passengers directly between O-D pairs of the ends in the subway network branches in a shorter period of time and improve passenger travel efficiency. To maximize the number of additional customers that can be served, a mixed integer programming model is established based on a temporal-spatial network to improve this issue. Numerical experiments validate the effectiveness and robustness of the proposed strategy, and some management suggestions for rail transit operation agencies are also discussed.

58 - Seawater Desalination: Exploring Alternative Water Resources and Addressing Water Scarcity through Discrete Choice Experiment in South Korea

Suk-Hoon Lee, Kyunghee University, Seoul, Korea, Republic of, Hyunhong Choi

Rising temperatures and climate change-induced precipitation variability are reducing the availability of freshwater resources. In response, seawater desalination plants are expanding globally, including in China, Spain, and the Middle East. Meanwhile, South Korea, with its manufacturing industry at its core, faces critical industrial water supply challenges, leading to shortages in household drinking water.

To address the conflict between industrial and domestic water scarcity, dam facilities have played a primary role in South Korea. However, with decreasing freshwater availability, solely relying on dams has become insufficient. Therefore, there's increasing consideration in South Korea to expand seawater desalination facilities. Public and corporate opinions on this transition are crucial, yet proper surveys regarding their views have not been conducted. Hence, this study comprehensively evaluates public and corporate preferences and acceptance through a Discrete Choice Experiment (DCE) survey targeting 1,000 citizens and 409 coastal businesses, predicting market share for seawater desalination facilities using choice simulation.

According to the research findings, seawater desalination facilities utilizing renewable energy significantly improve alternative choice probabilities and acceptance compared to traditional inland dams. This study emphasizes the importance of seawater desalination facilities in addressing domestic industrial and drinking water scarcity issues and suggests that integrating renewable energy facilities enhances their effectiveness.

Such research is perceived as an essential strategy involving the nation, businesses, and households. The success of seawater desalination projects will lead to stable industrial water supply for businesses, promoting national development and economic growth while ensuring future generations' access to drinking water.

59 - The effect of patient-to-nurse ratios and patient outcomes: Intensive Care Units versus Medical and Surgical Units

MARBELLY DAVILA, Tampa General Hospital, Tampa, FL, United States

The issue of nurse-to-patient ratios has gained significant attention due to the inadequate staffing levels in healthcare settings across the country. In large academic hospitals, nurses are frequently burdened with caring for more than expected patients at a time, amidst a shortage of nursing assistants. It is imperative to recognize that evidence-based practice strongly supports the notion that lower nurse-to-patient ratios lead to improved patient outcomes. However, it is essential to acknowledge that current policies often do not align with this best practice.

60 - Proposing a Time-to-Event Analysis for Assessing Supply Chain Disruption Severity and Occurrence Using a Competing Risk Model

Ronald Rineer, Widener University, Chester, PA, United States, Amin Keramati

Supply chain disruptions, ranging from minor delays to complete cessation of operations, can significantly impact a business's operations, leading to increased costs and revenue loss. To mitigate these effects, it is essential to predict the likelihood, timing, and severity of such disruptions. Previous models mostly used time series predictive approaches to assess supply chain disruptions. However, in this study, we consider supply chain disruptions as unexpected events, making their prediction more complex than analyzing repetitive patterns of event occurrence. To address this gap, this study treats supply chain disruptions as a time-to-event problem and employs a type of survival analysis known as the Competing Risk Model (CRM). The proposed predictive model can assist supply chain and logistics decision-makers in redefining committed delivery times, determining which products to invest in, and making informed decisions about suppliers, transportation modes, and other logistics considerations.

61 - The Benefit of Interpersonal Uncertainty: On the Design of Reward Allocation Mechanism in Online Referral Reward Programs

Huijie Jin, Renmin University of China, Beijing, China, People's Republic of, Shouwang Lu, Kanliang Wang

The online referral reward program (ORRP) is a topline marketing strategy for businesses to grow the customer base, in which firms promote their business via existing customers' online social networks and reward both the “referrer” and the “referee”. As a key component of the ORRP, rewards play an indispensable role in motivating users to engage in referral behaviors. However, inappropriate reward allocation design might harm the relationship between referrers and referees, ultimately reducing referring. We seek to explore the effect of the uncertain reward allocation, a novel reward allocation design in ORRP practice. Compared to the certain allocation, in the uncertain allocation context, both referrer and referee are only aware of the total referral reward, with the respective allocated reward amounts ex-ante unknown. Given that ORRP participants primarily perceive and evaluate rewards through comparisons, this research proposes the impact of uncertain reward allocation on referrers' referral intentions from the theoretical perspective of social comparison. The proposed hypotheses were investigated by three experiments. The results indicate that the uncertain (vs. certain) reward allocation of ORRP increases the referrers' referral intention by reducing their psychological barriers associated with explicitly comparing rewards with the referee. Furthermore, compared to large social distances (acquaintances), when the referrer considers making referrals to small social distances (friends) where they generally have low psychological concerns, the positive impact of uncertain reward allocation on the referral intention is weakened. The research findings provide new insights for both theory and practice.

62 - Tracing Sedimentary Origins in Multivariate Geochronology via Constrained Tensor Factorization

Naomi Graham, University of British Columbia, Vancouver, BC, Canada, Nicholas Richardson, Michael Friedlander, Joel Saylor

Tracing the origins of sediment over time can increase our understanding of the formation of geological structures as well as help to predict the presence of precious minerals. Detrital zircon grains serve as discrete carriers of information as their durability allows them to maintain elemental compositions over large time scales. These elemental concentrations can be measured and used to study sedimentary data, namely in relation to how it is comprised of sediments travelling from different sources. Current quantitative methods for analyzing sedimentary data are limited to studying up to two elemental features simultaneously, thus limiting their applications. We develop a model based in tensor decomposition and implement a coordinate descent algorithm for processing sedimentary data using an arbitrary number of features, while scaling linearly in the number of features. Our experiments indicate that a larger number of features increases our accuracy of identifying sedimentary sources, and we test our method on real data, achieving over 85% accuracy.

63 - Regression and Time Series Mixture Approaches to Predict Resilience

Priscila Silva, University of Massachusetts Dartmouth, Dartmouth, MA, United States, Gaspard Baye, Mindy Hotchkiss, Gokhan Kul, Nathaniel Bastian, Lance Fiondella

Resilience engineering is the ability to build and sustain a system that can deal effectively with disruptive events. Previous resilience engineering research focuses on metrics to quantify resilience and models to characterize system performance. However, resilience metrics are normally computed after disruptions have occurred and existing models lack the ability to predict one or more shocks and subsequent recoveries. To address these limitations, this research presents three alternative approaches to model system resilience with statistical techniques based on (i) regression, (ii) time series, and (iii) a combination of both to characterize system performance under multiple shocks and stresses of different intensity and duration, provide structure for planning tests to assess system resilience against particular degradations and guide data collection necessary to conduct tests effectively. These modeling approaches are general and can be applied to systems in multiple domains. A historical data set on job losses during the 1980 recessions in the United States is used to assess the predictive accuracy of these approaches. Goodness-of-fit measures and confidence intervals are computed to assess how well the models perform on the data set considered. The results suggest that resilience models based on statistical methods such as multiple linear regression and multivariate time series models are capable of modeling and predicting resilience curves exhibiting multiple shocks and subsequent recoveries. However, models that combine regression and time series account for changes in performance due to current and time-delayed effects from disruptions most effectively, demonstrating superior performance in long-term predictions despite increased parametric complexity.

64 - Automation design to support multitasking performance

Jiaxin Li, University of Washington, Seattle, WA, United States, Zishu Ling, Karine Suryatna, Ji-Eun Kim

Automation, a complete or partial transfer of operators’ tasks to a machine or computer, has reportedly reduced operators’ workload, resulting in enhanced performance in high-demand multitasking environments. Yet, little is known about when and how to use automation in consideration of task-related factors associated with multitasking performance. In this empirical study, we aim to evaluate the effect of two task-related factors: sensory modality (intra- vs. cross-modality) and task priority (absent vs. present) on multitasking performance employing the Multi-Attribute Task Battery, a computer-based in-flight task designed to simulate multitasking environments. Our results show that automation improved multitasking performance by reducing reaction time and error counts in cross-modality conditions, while only reaction time improved in intra-modality conditions. Reduced eye fixation entropy in the cross-modality condition indicates that reduced attention diversion was associated with automation support. Assigning higher priority to the automated task led to faster detection of automation failure, regardless of modality. Our findings offer insights into designing automation taking into account contextual factors in various work settings where operators need to perform multiple tasks under high workload situations.

65 - Deciphering NFTs: Understanding Decision-Making with Interpretable Machine Learning

LANQING DU, Drexel University, Philadelphia, PA, United States, Tianyi Ren

Non-Fungible Tokens (NFTs) stand as distinctive digital assets, each bearing a singular digital identifier denoting ownership, facilitated through blockchain technology.
In December 2021, the digital artwork "Merge," crafted by the renowned artist Pak, commanded a staggering $91.8 million on the NFT platform Nifty Gateway. NFTs have garnered significant attention due to their record-breaking sales prices, underscoring their importance within the digital economy. Our aim is to elucidate the decision-making processes involved in NFT (Non-Fungible Token) transactions through the lens of an interpretable machine learning approach.

66 - Stochastic contextual bandits with graph feedback: from independence number to MAS number

Yuxiao Wen, New York University, New York, NY, United States, Yanjun Han, Zhengyuan Zhou

We consider contextual bandits with graph feedback, a class of interactive learning problems with richer structures than vanilla contextual bandits, where taking an action reveals the rewards for all neighboring actions in the feedback graph under all contexts. Unlike the multi-armed bandits setting where a growing literature has painted a near-complete understanding of graph feedback, much remains unexplored in the contextual bandits counterpart. In this paper, we make inroads into this inquiry by establishing a regret lower bound $\Omega(\sqrt{\beta_M(G) T})$, where M is the number of contexts, G is the feedback graph, and $\beta_M(G)$ is our proposed graph-theoretic quantity that characterizes the fundamental learning limit for this class of problems. Interestingly, $\beta_M(G)$ interpolates between $\alpha(G)$ (the independence number of the graph) and $\mathsf{m}(G)$ (the maximum acyclic subgraph (MAS) number of the graph) as the number of contexts M varies. We also provide algorithms that achieve near-optimal regret for important classes of context sequences and/or feedback graphs, such as transitively closed graphs that find applications in auctions and inventory control. In particular, with many contexts, our results show that the MAS number essentially characterizes the statistical complexity for contextual bandits, as opposed to the independence number in multi-armed bandits.

67 - Exploring equitable accessibility for a foodbank network

Rashik Intisar Siddiquee, North Carolina State University, RALEIGH, NC, United States, Julie Ivy, Lauren Davis, Osman Ozaltin, Dwight Lewis

Food insecurity is a significant humanitarian challenge, with 12.8% of U.S. households experiencing food insecurity in 2023. Food banks play a crucial role in addressing this issue with their wide network of partner agencies, but ensuring equitable distribution and accessibility remains a complex challenge. This study examines the transportation accessibility of a food bank network using Geographic Information Systems (GIS) and data visualization tools. The objective is to identify underserved areas by considering food-insecure populations and food bank agencies that lack adequate public transport access. This analysis can better inform food banks to make additional interventions, ensuring more effective and equitable service distribution.

68 - Near-Optimal Cost Function Approximation for Multi-Period Technician Routing and Scheduling

Kaiwen Li, Beijing Foreign Studies University, Beijing, China, People's Republic of, Xi CHEN

We study the problem of designing near-optimal formulation of cost function approximation (CFA) for multi-period technician routing and scheduling problem in an online learning setting. Prior works using CFA design the parameterized model based on domain knowledge and tune parameters in an offline setting without providing error analysis. However, when the impact of uncertainty is easy to recognize, we propose an approach to find a near-optimal formulation of CFA a prior. More precisely, we first build Kullback-Leibler ambiguity set of future cost incurred by certain action, then we use the relationship between regularization and robust optimization to construct an equivalent regularization term, which leads to the modified cost function in CFA. Next, the performance gap between our proposed approach and optimal policy is given by error analysis of state aggregation. Numerical experiment demonstrates that the proposed approach (without tuning parameters) performs close to optimal in various settings of urban delivery.

69 - Breaking or Exacerbating the Cocoon: Exploring the Impact of Personalized Recommendations on Users’ Information Cocoons

Molin Yang, Peking University, Beijing, China, People's Republic of, Lin Wang, Chong Wang, Jiayin Zhang, Sean Xu

Information cocoon refers to the phenomenon where people are selectively exposed to information consistent with their beliefs or preferences. Literature on social media platforms primarily attributes the formation of information cocoons to filtering algorithms. However, the significant role of individual self-selection has largely been overlooked. This study distinguishes between two primary processes involved in information cocoon formation: algorithmic filtering and individual self-selection. We define two types of cocoons: recommendation cocoons, arising from algorithmic filtering, and selection cocoons, stemming from the self-selection process. Collaborating with one of the most popular social media platforms in China, we investigate how these two cocoons co-evolve and form the ultimate cocoons. This platform introduced a function that allows users to disable personalized recommendations. Our analysis delves into the impact of disabling personalized recommendations on recommendation cocoons, selection cocoons, and the ultimate cocoons.

Our findings reveal that disabling personalized recommendations directly reduces the extent of users’ recommendation cocoons while increasing the extent of selection cocoons. A possible explanation is that disabling personalized recommendations reduces the effectiveness of recommended content, which may lead to information overload and increased search costs, therefore exacerbating information narrowing in self-selection process. For users with higher degree of information cocoons, the extent of ultimate cocoons is reduced after disabling personalized recommendations. However, for users with lower degree of cocoons, disabling personalized recommendations increases the extent of ultimate cocoons, indicating the potential of recommender systems to break cocoons. Moreover, across all users, disabling personalized recommendations reduces engagement, highlighting the importance of personalized recommendations.

70 - A Reinforcement Learning-Based Model for Venture Capital Financing Rounds Investment

Guanrou Deng, University College London, London, United Kingdom

Investing in Venture Capital (VC) staged financing involves high risks and price volatility. Existing theoretical methods based on real options theory often simplify uncertainties and lack risk-averse considerations. Additionally, industry practitioners' decisions often involve personal biases and may not be analytically optimal. This paper introduces a novel reinforcement learning-based method to enhance venture capitalists' decision-making. Our policy-based agents are trained to leverage high payoffs and risks throughout the VC rounds of the financing period, which can allocate investments in proper rounds and identify optimal candidates to exit with the aim of maximising financial returns. Using Pitchbook data, we propose suitable exit multiples for different industries, and on average recommend an exit multiple of 10. Additionally, we find that prioritizing higher-performing companies enhances payoffs, though excessive selectivity may limit exits. These findings underscore the potential of reinforcement learning in optimizing VC-staged financing strategies, providing insights for informed decision-making in staged financing investments.

71 - Mixed-Effect Models for Assessing Variable Speed Limit (VSL) Signs

Julio Cesar Jurado, University of Maine, Orono, ME, United States, Mohammad Razaur Rahman Shaon, Niloufar Shirani, John Ivan, Naveen Eluru, Ali Shirazi

This study aims to assess driver responses to Variable Speed Limit (VSL) signs using mixed-effect linear and logistic regression models. The mixed effect models account for the impact of the VSL location and repeated observations. The data collected from the Vehicle Speed Sensor (VSS), and Road Weather Information Systems (RWIS) along Interstate 80 in Utah (between Salt Lake City and Kimball Junction) were consolidated to create a uniform dataset for analysis. This dataset covers a stretch of 14 miles of highway, from November 2021 to April 2023. First, a linear mixed-effect model is constructed to analyze the relationship between the speed downstream of the VSL sign and several factors, including upstream speed, traffic count per lane, weather conditions, and dummy variables indicating the time of day (morning peak hours, evening peak hours, and off-peak), roadway surface conditions (Frost-Slushy or Ice-Snow versus Dry-Damp-Wet), and the VSL activation. Second, a binary response model is developed using mixed-effect logistic regression to examine compliance with the speed limit when the VSL sign is activated. Our results show that these signs are effective in speed reduction during adverse weather conditions. The findings of this study provide valuable insights for decision-makers, demonstrating the real-world applications of VSL signs during adverse weather conditions.

72 - Enhancing Sample Diversity in Deep Convolutional GANs Using a Bayesian Framework

Mahsa Valizadeh, Texas A&M University, College Station, TX, United States, Rui Tuo, James Caverlee

Generative Adversarial Networks (GANs) are proficient at creating artificial data, but the persistent challenge of mode collapse during the optimization process necessitates innovative solutions. In our endeavor to address mode collapse in Deep Convolutional Generative

Adversarial Networks (DCGAN) and foster greater sample diversity, we introduce a Bayesian framework applied to DCGAN, referred to as Bayesian DCGAN. This framework involves the integration of a weight distribution within the network, achieved through the application of the Bayes by Backprop method. This method employs a mean-field variational inference approach to approximate the posterior distributions of weights. In addition to the incorporation of a weight distribution, we put forth a mathematical approach to quantify the diversity present in the samples generated by Bayesian DCGAN, contrasting it with the output of conventional DCGAN. Our experimental results showcase that Bayesian DCGAN generates more diverse samples compared to its conventional counterpart, thereby significantly reducing uncertainty in neural networks. This enhancement in diversity is pivotal for creating robust and adaptable models, particularly in scenarios where a broader spectrum of data representations is essential for effective learning and generalization.

73 - Smart Cities, Happier Lives? Big Data and Deep Learning Insights in China

Hongyu Zhang, Tsinghua University, Beijing, China, People's Republic of, Yidi WANG, Hongyu Liu

As sustainable urban development prioritizes improving residents' quality of life, smart cities, empowered by advancements in information technology, offer comprehensive management and service platforms that could enhance living conditions. However, the broader impacts and definitions of smart cities are often unclear due to limited data availability. This study addresses these gaps by analyzing big data from 300,000 smart city government procurement projects to categorize and quantify smart city investments. Utilizing deep learning techniques, specifically a RoBERTa fine-tuning classification model trained on manually labeled data, this research explores the correlations between different types and intensities of smart city investments and life satisfaction, considering the influence of political dynamics in China. The findings reveal a non-linear positive relationship between the intensity of smart city investments and residents' life satisfaction, with varying impacts depending on the type of investment. Investments targeting residents' livelihood services demonstrate the most substantial benefits, whereas infrastructure-focused projects yield less immediate improvements. Additionally, the analysis highlights how the career trajectories of government leaders can disrupt the alignment of long-term smart city investments, undermining their effectiveness and the welfare of the residents. This study not only advances understanding of smart cities through the lens of big data and deep learning but also offers critical insights for shaping future urban policies for sustainability.

74 - A Clinical Language Model to Detect Acute Traumatic Intracranial Injury from Free-Text Radiology Reports

Hayley Falk, University of Michigan, Ann Arbor, MI, United States

In acute traumatic brain injury (TBI), free-text brain computed tomography (CT) radiology reports contain detailed information regarding injury severity. This information is critically important for guiding clinical decision making and informing prognosis. However, the information contained within the radiology reports is unstructured and manual extraction of the relevant information is error-prone, difficult to scale, and requires significant resources (e.g., domain expertise, time, financial). Therefore, we developed and validated a clinical language model using radiology reports from 1531 emergency department (ED) patients evaluated for acute TBI. We developed our model using supervised finetuning of a clinical knowledge enriched large language model (LLM). The task was formulated as a binary text classification task (presence or absence of acute traumatic intracranial injury) and annotations were completed by a board certified neuroradiologist. We used a stratified 10 fold cross validation approach and selected the optimal hyperparameters based on performance in the validation set. Our final model demonstrated strong performance and had an area under the receiver operating characteristic curve (AUROC) of 0.99 and an F1 score of 0.90. Our results suggest that this is a promising approach that may help mitigate some of the limitations inherent in manual chart review. Additional analyses are needed to examine the generalizability of our model and to determine whether our model can be adapted to new tasks through finetuning.

75 - ADVANCED PREDICTIVE MODELS FOR DIABETIC RETINOPATHY: A MULTI-LAYERED ENSEMBLE APPROACH WITH HYPERPARAMETER TUNING

Mahyar Mahmoudi, Oklahoma State University, Stillwater, OK, United States

Diabetes is a prevalent chronic condition that often leads to severe complications such as diabetic retinopathy (DR), the leading cause of vision impairment in diabetic individuals. Our study analyzed data from over 1.4 million individuals to develop a predictive model for DR using demographic and laboratory data. This approach circumvents the need for specialized ophthalmic equipment and expertise. We employed a novel multilayered ensembling technique involving five algorithms: Linear Discriminant Analysis, Passive Aggressive Classifier, LinearSVC, Logistic Regression, and SGDClassifier. Through stack generalization, each model was fine-tuned across various hyperparameters to optimize performance metrics like precision, recall, and accuracy. The outputs from each algorithm were then aggregated through the second layer of ensembling, enhancing the interpretability and efficiency of the predictions. Our findings indicate that variables such as diabetic neuropathy, serum creatinine, blood urea nitrogen, glucose, and hematocrit are crucial for detecting DR. This insight allows for more targeted and efficient early screening, which could significantly improve intervention strategies. Our multilayered ensembling method, simpler and more interpretable than deep learning approaches, demonstrates how advanced analytics can be effectively integrated into clinical practices, setting the stage for further refinement and real-world application. This model offers a robust alternative for medical diagnostics, reducing computational demands while maintaining competitive accuracy.

76 - Knowledge Distillation-based Incremental Learning with Integration of Transformer and Its Applications in Online Process Monitoring

Boris Oskolkov, Oklahoma State University, Stillwater, OK, United States, Zhangyue Shi

The pursuit of ever-larger models for incremental learning raises concerns regarding computational cost and environmental sustainability. This work proposes a novel approach that prioritizes efficiency without compromising performance. We introduce a lightweight framework leveraging the knowledge preservation capabilities of transformers to facilitate effective yet sustainable incremental learning. The proposed method employs a compact, pre-trained transformer model as a "teacher." This teacher, despite its relatively smaller size, retains a wealth of knowledge from its initial training. Instead of directly participating in the computationally expensive incremental learning process, the teacher imparts its knowledge onto a smaller, more adaptable student model. This distilled knowledge acts as a stabilizing force, mitigating catastrophic forgetting as the student learns from new data. By offloading long-term memory onto the static teacher, the student model can be kept lightweight and computationally efficient during incremental training. The experimental results demonstrate that this approach achieves competitive performance on simulated dataset while significantly reducing computational demands compared to conventional methods. This work highlights a pathway towards incremental learning by combining the power of transformers with a focus on resource efficiency for future application in manufacturing.

77 - Complete formulation of the capability curve of synchronous generators in OPF models

Rafael Zarate-Minano, University of Castilla - La Mancha, Almadén, Spain, Miguel Carrion

This work addresses the limitations in the current modeling of synchronous generators in optimization problems, where power production limits are often simplified with independent maximum and minimum bounds on active and reactive power outputs. The study proposes a more comprehensive model that includes the complete capability curve of the machine, accurately representing its stability limits. The effectiveness of this model is validated through a maximum loading condition problem, using a homotopy modeling technique to push the generator to its operating limits. This approach offers a more precise representation of generator operation, enhancing the accuracy of optimal power flow solutions.

78 - Connected Vehicles Feature Analytics using Generative AI and Intelligent Sampling

Oleg Gusikhin, Ford Global Data Insight & Analytics, Dearborn, MI, United States, Omar Makke, Syam Chand, Vamsee Batchu, Svidenko Vicky

The increase of connected vehicles has led itself to having an unsurmountable amount of data. Analyzing connected features can become computationally expensive, or even impossible. This necessitates the need for a system that assists in properly sampling the signals using either classical methods or more recent Intelligent Sampling methods. In many cases selecting the correct vehicle signals and a proper sample for the given feature is not trivial. By leveraging Generative AI, this complexity can be efficiently managed. This is due to the ability of Generative AI to process design documents and combine them with general knowledge, to guide the users in collecting the proper data. This system has wide applications such as in engineering, design, and marketing.

79 - Tensor-Based Multi-modal Data Fusion

Mason Li, Chapman University, Orange, CA, United States

Machine learning is widely adopted but often requires substantial computational resources. This research addresses the need for more efficient models by exploring tensor decomposition as a method to reduce computational load. We aim to perform tensor decomposition on convolutional layer weights, therefore reducing dimensionality and enhancing computational efficiency. We aim to achieve this efficiency without sacrificing accuracy. The proposed model integrates multi-modal inputs, combining X-ray imagery and patient demographics, to perform length-of-stay regression and mortality classification. Experimental results indicate a significant reduction in model parameters and inference time, demonstrating the potential of tensor decomposition in real-time clinical applications.

80 - Stochastic Approximation with Delayed Updates: Finite-Time Rates under Markovian Sampling

Arman Adibi, Princeton University, Princeton, NJ, United States, Nicolò Dal Fabbro, Luca Schenato, Sanjeev R. Kulkarni, H. Vincent Poor, George J. Pappas, Hamed Hassani, Aritra Mitra

Motivated by applications in large-scale and multi-agent reinforcement learning, we study the non-asymptotic performance of stochastic approximation (SA) schemes with delayed updates under Markovian sampling. While the effect of delays has been extensively studied for optimization, the manner in which they interact with the underlying Markov process to shape the finite-time performance of SA remains poorly understood. In this context, our first main contribution is to show that under time-varying bounded delays, the delayed SA update rule guarantees exponentially fast convergence of the \emph{last iterate} to a ball around the SA operator's fixed point. Notably, our bound is \emph{tight} in its dependence on both the maximum delay τ_{max}, and the mixing time τ_{mix}. To achieve this tight bound, we develop a novel inductive proof technique that, unlike various existing delayed-optimization analyses, relies on establishing uniform boundedness of the iterates. As such, our proof may be of independent interest. Next, to mitigate the impact of the maximum delay on the convergence rate, we provide the first finite-time analysis of a delay-adaptive SA scheme under Markovian sampling. In particular, we show that the exponent of convergence of this scheme gets scaled down by τ_{avg}, as opposed to τ_{max} for the vanilla delayed SA rule; here, τ_{avg} denotes the average delay across all iterations. Moreover, the adaptive scheme requires no prior knowledge of the delay sequence for step-size tuning.

81 - Feature Subset Selection Framework for Endometriosis Diagnostics

Carolina Arroyo-Roldan, University of Puerto Rico - Mayaguez, Anasco, PR, United States, Wandaliz Torres-Garcia

Endometriosis is a gynecological disease characterized by the growth of endometrial tissue outside of the uterus. It is known to cause pelvic pain, dyspareunia, and dysmenorrhea, and has been linked to infertility. The diagnosis of endometriosis takes various years, partly because the gold standard for diagnosis is an invasive laparoscopic surgery which poses both a physical and financial burden on women. Previous research studies to identify potential molecular biomarkers face statistical and computational limitations due to small sample sizes with high dimensionality. Additionally, there is no standardization regarding factors that should be considered when analyzing the gene expression of endometrial tissue, such as disease severity and the phase of the menstrual cycle. Though experiments with large sample sizes are not available, many new and upcoming experiments are being generated to bridge the gap in understanding the biological mechanisms of this largely uncharacterized disease affecting many women worldwide. Therefore, we propose the development of a consensus framework to extract the best feature subsets to discriminate between endometriosis patients versus women without the disease using gene expression of endometrial tissue, as well as clinical information such as the stage of the menstrual cycle. This framework aims to identify and rank feature subsets using (1) optimization-based search algorithms coupled with classification methods and (2) queries of functional processes from gene-annotated databases across multiple publicly available experiments from the Gene Expression Omnibus repository to evaluate consensus.

82 - When Are Combinations of Humans and AI Useful?

Michelle Vaccaro, MIT, Cambridge, MA, United States, Abdullah Almaatouq, Thomas Malone

Inspired by the increasing use of AI to augment humans, researchers have studied human-AI systems involving different tasks, systems, and populations. Despite such a large body of work, we lack a broad conceptual understanding of when combinations of humans and AI are better than either alone. Here, we addressed this question by conducting a meta-analysis of over 100 recent experimental studies reporting over 300 effect sizes. First, we found that, on average, human-AI combinations performed significantly worse than the best of humans or AI alone. Second, we found performance losses in tasks that involved making decisions and significantly greater gains in tasks that involved creating content. Finally, when humans outperformed AI alone, we found performance gains in the combination, but when the AI outperformed humans alone we found losses. These findings highlight the heterogeneity of the effects of human-AI collaboration and point to promising avenues for improving human-AI systems.

83 - Ensuring Fair Prediction with Group-Aware Priors

Tim G. J. Rudner, New York University, New York, NY, United States

We consider the problem of training neural network models that make fair predictions in settings where certain subpopulations, or groups, are highly underrepresented in the training data. To address this problem, we develop a family of group-aware prior distributions over neural network parameters explicitly designed to favor models that are fair. Priors in this family assign high probability density to parameter values that induce models that are robust across groups. We show how to construct group-aware priors and derive a tractable optimization objective that incorporates them into model training. We demonstrate empirically that training with group-aware priors yields fair models on several challenging prediction tasks. Our evaluation shows that group-aware priors outperform state-of-the-art methods on vision and language classification tasks while only requiring access to a small amount of group information. Group aware-priors are conceptually simple, complement alternative approaches for fair prediction, and allow harnessing Bayesian inference to further improve model robustness.

84 - Enhanced Prediction of Hospital Readmission in Heart Failure Patients: Model Validation and Methodological Improvements Using Sepsis Patient Data

Sonia Jahangiri, Rochester Institute of Technology, Rochester, NY, United States, Nasibeh Azadeh-Fard

Predicting hospital readmissions is a major challenge in healthcare, particularly for heart failure patients, due to the complex interplay of clinical and demographic factors. Reducing readmission rates is crucial for improving patient outcomes and minimizing healthcare costs. This study focuses on developing a predictive model for hospital readmission in heart failure patients, which is then validated and refined using data from sepsis patients. By employing advanced machine learning techniques and various validation methods, the research enhances the model's accuracy and robustness. The process includes incorporating diverse patient variables, performing cross-validation, bootstrapping, and holdout methods, and fine-tuning hyperparameters. The results show significant improvements in prediction accuracy, demonstrating the effectiveness of these methodological enhancements for predicting hospital readmissions across different patient populations.

85 - Predicting Hospital-Acquired Conditions Among Children with Multiple Chronic Conditions: Individual Disparity and Structural Racism Effects

Ajit Appari, Northeastern University, Boston, MA, United States

Hospital-acquired conditions (HACs) are mostly preventable risk to the safety of hospitalized children, especially with multimorbidity that can result in elevated mortality & morbidity, longer hospitalization, and higher cost of care. Machine learning modeling is becoming an important tool for delivering equitable health care and ensuring better health outcomes.

In my study, I focused on building prediction model for composite risk of multiple hospital acquired conditions (MHAC) among chronically ill children having different multimorbidity patterns adjusting for individual disparity factors {age, sex, race/ethnicity, insurance}, structural racism factors {Neighborhood Disadvantage (e.g., poverty, unemployment, households on state income support, and higher proportion of Blacks), Neighborhood Affluence (highly educated, high income population) and Hispanic & Immigrant Concentration}, and provider factors (hospital characteristics).

I used administrative data on more than 700,000 pediatric hospitalizations (age: up to 17years) during 2015Q4-2016Q4 across 642 hospitals in Texas. Patients were coded for 69 chronic condition categories (reflecting 4400+ ICD10 diagnosis-and-procedure codes) and grouped into 10 different multimorbidity patterns using unsupervised learning method- Block Clustering with Bernoulli Mixture Models that leverages block classification expectation maximization algorithm. Subsequently, patients were coded for multimorbidity pattern complexity (number of chronic conditions from each pattern) to quantify potential complexity of care needs for presence of such patterns. Finally, I build and evaluate composite risk prediction model for MHACs using Zero-Inflated Negative Binomial Regression, and XGBoost Hurdle Model accounting for multimorbidity pattern complexity, individual disparity factors, structural racism factors, and provider factors.

86 - On the Adversarial Robustness of Benjamini Hochberg

Louis Chen, Naval Postgraduate School, Monterey, CA, United States, Roberto Szechtman

The Benjamini-Hochberg (BH) procedure is widely used to control the false detection rate (FDR) in (large-scale) multiple testing. Applications of this control abound in drug discovery, forensics, candidate screening, anomaly detection, and, in particular, machine learning, ranging from nonparametric outlier detection to out-of-distribution detection, and one-class classification methods. Considering this control is relied upon in some critical (safety/security) contexts, we investigate its adversarial robustness. More precisely, we study under what conditions BH does and does not exhibit adversarial robustness, we present a class of simple and easily implementable adversarial data-perturbation algorithms, and we perform computational experiments. With our algorithms, we demonstrate that it is possible for BH's control to be significantly broken with relatively few (even just one) test score perturbation(s), and provide non-asymptotic guarantees on the expected adversarial-adjustment to FDR. Our technical analysis involves a combinatorial reframing of the BH procedure as a ``balls into bins'' process, and drawing a connection to generalized ballot problems to facilitate an information-theoretic approach for deriving non-asymptotic lower bounds.

87 - Estimation of Waiting Time for Employment-Based Permanent Residency in the US: A Machine Learning Framework

Meghan Aerick, California Polytechnic State University, San Luis Obispo, CA, United States, Puneet Agarwal

The United States receives a substantial number of green card applications from individuals seeking permanent residency each year. The processing time for these applications varies significantly based on a variety of factors, causing much uncertainty and anxiety for applicants. This study presents the development and implementation of a predictive application using machine learning (ML) models to estimate U.S. employment-based green card waiting times. Our model leverages historical data from the United States Department of Labor and current processing times from the United States Department of State Bureau of Consular Affairs to predict waiting times. The predictive tool, developed using machine learning algorithms including Decision Trees, Random Forest, AdaBoost, and XGBoost, is hosted on a user-friendly Streamlit application. It offers an intuitive user interface that guides users through data input, model selection, and visualization of predictions and explanatory analyses. The best-performing ML model achieved an F1-score macro of 0.92 and an accuracy of almost 93%, indicating high reliability. Aimed at assisting applicants in effectively planning their futures, this application simplifies complex ML predictions into accessible insights, thereby democratizing access to predictive analytics. Additionally, it includes pages for model performance metrics, feature importance analyses, and general data insights to educate users about the underlying processes and factors influencing wait time predictions. This poster outlines the development process of the application, offers a thorough descriptive and predictive analysis of the data, and highlights the key functionalities of the Streamlit-based decision analysis tool.

88 - An Analytical Model for Assessing Factors Influential in Long- Versus Short-Term Survival after Kidney Transplantation

Mostafa Amini, California State University Long Beach, Long Beach, CA, United States, Aayush Doshi, Hassan Ahmed Qureshi

The significant gap between the availability of organs and the number of patients on transplant waitlists necessitates an equitable and optimized organ allocation system. This system should aim to minimize waitlist mortality and enhance the benefits of transplantation, such as improved survival time and quality of life. This project introduces the development of an explanatory analytics framework designed to uncover the key factors potentially influencing long-term survival following kidney transplantation. Utilizing data from the United Network for Organ Sharing registry, which includes over 100 thousand unique kidney transplant patients in the US since 1987, the framework will employ a variety of machine learning (ML) algorithms for initial classification. The optimal algorithm will subsequently be supplemented with advanced explainable artificial intelligence techniques to allow for comprehensive model explanation and interpretation. A crucial component of our approach involves extensive feature engineering to discover more nuanced relationships among variables, including interaction effects and non-linearities, which are often overlooked in traditional analyses. This framework aims to identify and analyze potentially critical but underexplored factors. By offering a detailed examination of these factors, the proposed framework seeks to significantly enhance the kidney allocation process, potentially leading to improved outcomes for recipients. The insights generated from this research could provide invaluable guidance for medical scholars and practitioners, informing future enhancements to organ transplantation management strategies.

90 - Fixing the Broken Rung: On The Interactions of Allyship and Microaggressions for Women ICT Professionals

Shiyi Wang, Nanyang Technological University , Singapore, Singapore, Anandasivam Gopal, Sierin Lim

The persistent underrepresentation of women at senior and mid-career levels in the information systems (IS) sector—the "*broken rung*" problem—is extensively recognized. However, the understanding of its underlying causes and effective solutions is still lacking. Through collaboration with four multinational technology firms in Singapore, we recruit high-quality participants to explore the broken rung problem. Our findings reveal that microaggressions substantially diminish the psychological safety of mid-career women in the ICT field, reducing promotability perceptions and lower job satisfaction. Concurrently, allyship is identified as a potential mitigating factor, buffering the adverse effects of microaggressions by bolstering mid-career women's psychological safety. Interestingly, our results indicate that men overlook these dynamics, which suggests that "privilege blindness" exists in the ICT field. This calls for further efforts to enhance organizational awareness and create actionable strategies such as cultivating allyship within organizations to solve the broken rung problem.

91 - Revenue Management, Loyalty Program, and Online Travel Agency in Hotel Chains

Shuxian Xiao, State University of New York-Buffalo, Buffalo, NY, United States, Charles Wang, Mingcheng Wei

We study the hotel's optimal designs of the premium-status loyalty programs in two alternative channels selling rooms to strategic consumers: direct channel versus dual channels. In the direct-channel scenario, the hotel sells rooms directly to consumers, while in the dual-channel scenario, the hotel sells rooms both in its direct channel and through an independent channel owned by an online travel agency (OTA). Note that the hotel excludes OTA bookings from the premium-status rewards. We characterize the consumers' purchasing behavior as well as the hotel's optimal pricing and premium-status loyalty program design decisions in both direct-channel and dual-channel scenarios. In the direct-channel model, we relax the restriction of the qualification threshold in a prior work, derive a complete spectrum of consumer segmentation, and find that the hotel can significantly improve its profit by offering a low qualification threshold to make its premium status more achievable by customers. We also numerically compare the hotel's optimal decisions and profits between direct- and dual-channel scenarios, and we find that the hotel can improves its profit by up to 2.57% with the dual-channel model.

92 - Return Based Risk Parity

Viraat Singh, Columbia University, New York City, NY, United States

- Traditional risk parity provides a way of diversifying a portfolio while preventing excessive risk concentration, providing a way to construct portfolios good risk diversification. The volatility of risk parity portfolio lies somewhere between the minimum variance and the 1/n portfolio that has extensively been studied. However, risk parity portfolios or equal risk contribution portfolios (ERC) depend only on the covariance matrix of the asset universe and is agnostic to the returns of the assets. Efforts have been made to incorporate returns into risk parity and other risk budgeting methods, but these methodologies produce non-convex optimization problems that are difficult to solve or relaxations that have no guarantees on the portfolio obtained. This not only makes them hard to solve using numerical heuristics, but these heuristics give no guarantees on the risk diversification. In this paper, we adopt the principle of diversifying risk contributions to improve returns, by satisfying approximate risk parity whilst providing bounds on risk spread and taking returns into account. Mathematically, we provide algorithms (RAH, RAC, AERC), that bound the gap between the risk contributions or risk spread (RS) and allows profitable assets to contribute more to a portfolio than would be allowed through regular risk parity.

93 - A fast nonparametric process monitoring scheme in fully decentralized systems

Jiahui Zhang, University of Wisconsin- Madison, Madison, WI, United States, Ziqian Zheng, Kaibo Liu

Along with the development of big-data technologies, multi-sensor systems have become more and more prevalent in manufacturing industries. To ensure the production process is in normal condition, process monitoring algorithms have been applied to multi-sensor systems, in which a central server aggregates information from all sensors and make decisions about the process condition. This kind of centralized systems, however, highly relies on the central server and cannot protect the privacy of the sensor data. This leads to the need of decentralized systems, in which each edge device can communicate with each other and individually raise the system alarm. The goal of this paper is to develop a fully decentralized process monitoring scheme for fast online change detection. Our monitoring framework includes the design of a nonparametric device-level statistic, a smart strategy for inter-device communication, and an alarm-raising mechanism which responds quickly to mean shifts and is resilient to partial function failures of the devices. The proposed method enjoys good theoretical properties which enables its effectiveness and efficiency. The outperformance of the proposed method is shown in numerical experiments and a case study.

94 - Integration of renewable energy sources in electricity price formation in Pakistan through Merit-Order Effect.

SULEMAN MUHAMMAD, Xi'an Jiaotong University, Xi'an, China, People's Republic of

We examine how the integration of renewable energy sources has impact extensive electricity price in Pakistan. Our main concern is scheduling of electricity power plants to full fil demands via merit order effect. Through merit order effect the most cost-effective indicator is given into service, then the next one most cost effective, and so on. Renewable energy sources are preferred over conventional power plants more and more often due to their less operation costs. This is overall resulting in very cheaper production of electricity. In this article the merit order impact is analyzed using different data sets. This study uses daily data from 2010 to 2022 to view the merit order effect and SARMAX/EGARCH methods has been applied. if renewable energy sources have added to the country's electricity mix, the results explain that electricity price in Pakistan have fall down significantly. In some marketplaces the merit order effect appears stronger than others. The integration of renewable energy sources is affecting the electricity price formation in Pakistan with positive gesture, shown by this study. Utility rates are reducing through merit order effect, it shows progressive development for both consumers and entrepreneurs.

Keywords: Merit-Order effect, extensive electricity price, Renewable energy source, Exponential GARCH, Integration.

95 - A Linearized Branch Flow Model Considering Line Shunts for Radial Distribution Systems and its Application in Volt/VAr Control

Hanyang Lin, Tsinghua University, Beijing, China, People's Republic of, Ye Guo, Hongbin Sun

In the process of modernizing power distribution networks, one common change is the replacement of overhead lines with underground cables. Traditional linear power flow models, assuming zero shunt elements, become inadequate for modern distribution systems where line shunts are non-negligible. To address this, a linearization method for branch flow model considering line shunts (LBFS) is proposed. This model effectively captures the errors introduced by line shunts while maintaining a linear structure of π equivalent circuit. The model's utility is demonstrated in a two-stage stochastic Volt/VAr control framework to manage network voltages, with the uncertainty in active power generation from renewable energy sources.

96 - Designing Sparse Flexibility Structures: What to do when the Long Chain Fails?

David Corredor, Amazon, Boston, MA, United States, Ling Zhang

When designing sparse flexibility structures, the long chain is the best candidate. Theoretical results and empirical evidence show that the long chain structure can achieve a really good performance. This performance is compromised when assumptions presented in the literature are relaxed. The most restrictive assumption when designing flexibility structures in real-world applications is that the demands in the network are identically distributed. In this paper we intend to relax that assumption and consider arbitrary demand distributions. We leverage our methodology on existing results and build sparse flexibility structure with at most O(2n) and O(3n). The proposed structures show promising results by achieving almost optimal performance over several systems with arbitrary distributions.

97 - Predicting Kidney Transplant Success and Risk Factors

Guldasta Hajizade, The University of Tulsa, Collins College of Business, Tulsa, OK, United States, MD Talha Mohsin, Narges Salehin, Kazim Topuz

This study seeks to enhance kidney transplant outcomes by integrating machine learning (ML) techniques for predictive analytics. We employ feature selection algorithms, including RF-RFE (Random Forest Recursive Feature Elimination), DT-RFE (Decision Tree Recursive Feature Elimination), Lasso-RFE (Lasso Recursive Feature Elimination), and LR-RFE (Logistic Regression Recursive Feature Elimination), to identify critical predictors influencing transplant risk. These selected features inform the development of predictive models using Gradient Boosting, AdaBoost, and Artificial Neural Networks (ANNs). The performance of these models is evaluated using metrics such as accuracy, precision, recall, and Area Under the Curve (AUC), providing a comprehensive assessment of model accuracy and reliability. Furthermore, a feature importance analysis is conducted using SHapley Additive exPlanations (SHAP) to gain deeper insights into the factors significantly impacting transplant outcomes. Through a combination of careful feature selection, advanced modeling techniques, and thorough performance evaluation, this research aims to demonstrate the potential of ML to improve predictive accuracy in kidney transplantation. The application of these methodologies aspires to support better clinical decision-making and optimize patient outcomes in kidney transplantation procedures.

98 - Dynamic Interlay and Intraday Scheduling

Christos Zacharias, University of Miami, Coral Gables, FL, United States

The simultaneous consideration of appointment day (interday scheduling) and time of day (intraday scheduling) in dynamic scheduling decisions is a theoretical and practical problem that has remained open. We introduce a novel dynamic programming framework that incorporates jointly these scheduling decisions in two timescales. Our model is designed with the intention of bridging the two streams of literature on interday and intraday scheduling and to leverage their latest theoretical developments in tackling the joint problem. We establish theoretical connections between two recent studies by proving novel theoretical results in discrete convex analysis regarding constrained multimodular function minimization. Grounded on our theory, we develop a practically implementable and computationally tractable scheduling paradigm with performance guarantees. Numerical experiments demonstrate that the optimality gap is less than 1% for practical instances of the problem.

99 - Metamodel Initialization of a Search Algorithm for Repair Kit Application

Xueying Wang, Boston University, Boston, MA, United States, Liying Zhang, Ruthairut Wootisarn

The “repair kit problem” in inventory management involves optimizing the storage and delivery of parts that are used to repair replacement. This parts have interdependencies because the repair cannot be done until all the parts are available. This problem is particularly challenging because of the necessity of maintaining a balance between carrying costs and the risk of delaying the repair when one or more parts are unavailable. It exists in different industries, such as on-site repairs, submarines, space stations, and public transportation systems. This poster presents an approach to initialize an empirical stochastic branch-and-bound search algorithm using metamodels. The algorithm is included in a simheuristics framework, which combined with simulation and heuristics methods, to create periodic review inventory policies for parts in a warehouse that holds intermittent demand spare parts. The simulation mimics a repair kit, while the heuristic model calculates annual total cost slopes on a part-by-part basis. Initialization metamodels are derived from results that use a deterministic model to start the search. Metamodels of service level and order quantity were used to calculate the reorder points and order-up-to levels. This approach shows good accuracy when based on validation using a small repair kit.

100 - Research on the dynamic pre-sale strategy of fresh agricultural products considering supplier risk avoidance

yuxiu liang, southeast university, nanjing jiangsu province, China, People's Republic of, Lindu Zhao

Study the dynamic pre-sale strategy of fresh agricultural products when suppliers face planting risk (out of stock or overplanting) under the risk avoidance behavior of suppliers. The supply chain includes a supplier and a retailer. The supplier makes pre-sale decisions before planting and maturity, and the retailer makes sales decisions at the present sale stage. As strategic consumers, consumers decide which stage to buy according to their own utility maximization. Suppliers reduce the uncertainty of planting risk through pre-planting booking and pre-mature booking, while their risk aversion behavior is modeled through the CARA utility function. Retailers sell spot after the crops mature and optimize their sales strategies based on market demand and supply. In this paper, we analyze the optimal decision of supply chain members at each stage by establishing a multi-stage dynamic game model. The results show that the degree of risk aversion of suppliers and the uncertainty of market demand significantly affect their pre-sale quantity and price strategy. The numerical simulation shows that the higher the risk aversion, the lower the pre-sale quantity and the higher the price; as strategic consumers, the purchase time is affected by the price and available quantity. The research in this paper provides theoretical basis and practical guidance for the formulation of pre-sale strategy in the supply chain of fresh agricultural products

101 - Investment, partial irreversibility, and competition

Takashi Shibata, Tokyo Metropolitan University, Hachioji, Japan, Michi Nishihara, Yuan Tian

This study considers the optimal investment timing decision when the investment is not completely irreversible in a competitive market. As shown in Section 8.2 of Dixit and Pindyck (1994), the optimal investment decision is independent of the degree of competition when the investment is completely irreversible. This study finds that the optimal investment decision depends on the degree of competition when the investment is not completely irreversible. As an extreme case, if the competition is strongly intense, the firm never invests. The weaker the competition, the later the investment is exercised.

102 - Liquidity Identification and Trading Strategies at Dark Pools

Ruijing Yang, Stevens Institute of Technology, Hoboken, NJ, United States, Somayeh Moazeni, Ricardo Collado

To limit the adverse price impact of large orders and to realize more favorable prices, institutional investors may resort to Dark pools for block trading. However, due to the low execution rate of submitted limit orders, identifying liquidity patterns within dark pools is challenging. Building on massive data sets of 200 million orders for 7000 stocks across 43 venues, we investigate liquidity patterns in different dark pools and present several machine learning models to predict the likelihood of executing a submitted limit order within a specific dark pool at any time. Leveraging our feature selection results, we then develop trading strategies and assess their performance.

103 - Building Trust in Digital Twin Through Verification and Validation

Julia Bitencourt, Auburn University, Auburn, AL, United States, Gregory Harris

Digital Twin (DT) is a topic of growing interest, as it combines many technologies, such as Internet-of-Things, simulation, and machine learning to support decision-making in constantly changing environments. As the landscape of DT grows, it is fundamental to trust in all components of a DT, especially in manufacturing environments where DTs support decisions that directly impact cost and safety. Model trust is traditionally viewed from the perspective of Verification and Validation (V&V). In this context, establishing trust includes performing verification and validation activities to demonstrate how the V&V evidence can be applied to support the use of the DT. Even though the importance of V&V is extensively recognized as a fundamental step in developing DTs that can be trusted and used in real-world environments, very few works report implementing such activities. In this research, we uncover the status of V&V for digital twins. A methodology to support practitioners in developing trusted DTs is proposed. The methodology is applied to a case study in a fused deposition modeling additive machine to demonstrate how V&V can support the development of trusted DTs.

104 - Unveiling Shippers' Behavior through a Two-Sided Market Approach

Matthew Roorda, University of Toronto, Toronoto, ON, Canada, Azin Golrizkhatami, Chi-Guhn Lee

The online food delivery sector's rapid growth has been paralleled by the adoption of crowd shipping models, where delivery tasks are outsourced to individuals utilizing their idle transportation capacity. This paper presents a novel mathematical framework within the Markov Decision Process (MDP) to examine crowd shippers' behaviours, particularly focusing on decision-making processes in platforms like UberEATS. The study aims to identify optimal strategies for crowd shippers to maximize long-term earnings by determining which delivery offers to accept or reject. By implementing reinforcement learning algorithms, the findings suggest strategies that could enhance profitability for crowd shippers while providing insights into effective pricing policies for crowd shipping services. This research contributes to the broader discourse on optimizing gig economy operations, emphasizing the need for equitable compensation models.

105 - Multi-physics Guided Diffusion Models with Manufacturing Applications

Naichen Shi, University of Michigan, Ann Arbor, MI, United States, Hao Yan

We present a generic physics-data fusion generative model called MGDM that integrates multi-fidelity physics simulations with diffusion processes. MGDM categorizes multi-fidelity physics simulations into inexpensive and expensive simulations, depending on the computational costs. The inexpensive simulations, which can be obtained with low latency,
directly inject contextual information into diffusion models that effectively learn the conditional distribution through first-order optimization methods. Furthermore, when results from expensive simulations are available, MGDM refines the quality of generated samples via a guided diffusion process. This design separates the training of a denoising diffusion model from a physics conditional probability model, thus lending flexibility to practitioners.
MGDM builds on Bayesian probabilistic models and is equipped with a theoretical guarantee that provides upper bounds on the KL-divergence between the sample and underlying true distribution. We use a numerical simulation in fluid dynamics and a case study in laser-based metal powder deposition additive manufacturing to demonstrate how MGDM achieves a delicate balance between multi-fidelity physics knowledge and statistical
patterns from observations.

106 - Explainable AI in Transport Logistics: Enabling Smarter Decision-Making in Crash Mitigation

Shamkhal Mammadov, The University of Tulsa, Tulsa, OK, United States, Ismail Abdulrashid

Traffic accidents threaten global public safety and kill countless people each year. This poster presents an explainable artificial intelligence (XAI) artifact design that includes accident data for stakeholders and decision-makers. It proposes responsible, explanatory, and interpretable models using a systems-level taxonomy to categorize driver-related behaviors by injury severity, thereby contributing to transport logistics decision making. To determine and analyze risk factors influencing injury severity, we used advanced techniques such as data missing at random (MAR) with Bayesian dynamic conditional imputation for missing records, synthetic minority oversampling for data imbalance, and categorical boosting (CatBoost) with SHapley Additive exPlanations (SHAP). In addition, exploratory feature analysis revealed spatiotemporal elements that influence traffic accidents and injury severity. The second phase involved creating predictive models using XGBoost, RF, DNN, and fine-tuned parameters. Using the SHAP approach, we employed model-agnostic interpretation techniques to separate explanations from models. In the final phase, we provided an analysis and summary of the system-level taxonomy across feature categories. This involved classifying crash data into high-level causal factors using aggregate SHAP scores, illustrating how each risk factor contributes to different injury severity levels.

107 - Optimal Fulfillment and Transshipment Strategies in Omnichannel Retailing with Cross-Channel Returns

Yuchi Guo, University of Michigan- Dearborn, Dearborn, MI, United States, Armagan Bayram

To satisfy customer demand and maintain their competitiveness, retailers are exploring ways to exploit rising online sales, such as by investing in cross-channel strategies. Addressing the challenges in operating both the online and offline channels efficiently, in this study, we build a dynamic programming model and investigate optimal fulfillment and transshipment strategies in omnichannel retailing with cross-channel returns. With the objective of maximizing the total profit of the retailer, we investigate (i) from where to fulfill a home delivery order when it occurs, (ii) when and how to transship returns to balance inventory. In order to develop the optimal omnichannel fulfillment and transshipment policy, we accommodate the uncertainty in the customer demand, returns, and the product per unit cost of shipping and transshipping.

108 - Blame attribution when crowdsourced delivery fails: consumer perceptions in tetradic arrangements in online grocery shopping

Annibal Sodero, The Ohio State University, Columbus, OH, United States, Vince Castillo, Arianna Seghezzi, Chiara Siragusa, Angela Tumino

Crowdsourced delivery of online grocery purchases takes place in a tetradic arrangement, where a consumer purchases grocery made available by a retailer; a worker performs the physical tasks of picking the purchase order, selecting substitutes whenever necessary, and delivering the order, while a crowdsourcing platform coordinates the grocery purchase and the delivery, while exchanging information with both consumer and a worker. This study investigates to whom the consumer attributes blame whenever a failure occurs in a service encounter involving all four entities. Results shed light on design and execution of crowdsourced delivery of online grocery purchases.

Monday, October 21, 2:15 PM - 3:30 PM

MD01

Summit - 320

Scheduling and Its Contemporary Applications

Invited Session

Scheduling and Project Management

Chair: Zhixin Liu, University of Michigan-Dearborn, Dearborn, MI, United States

1 - Strengthening Priority Supply Chains: ***A Case of Candidate Supply Chains Selection Through Data Analytics Methods***

Charu Chandra, University of Michigan - Dearborn, Dearborn, MI, United States, Jorge Laboy-Bruno, Sam Le

In early 2023, the United States President's Management Agenda proclaimed a commitment to identify government-wide supply chains for prioritization due to diversity, economic, or national security risk. Among various objectives identified for this initiative were, (a) for federal procurement agencies identifying and giving attention to market segments within their own supplier base, where they have experienced challenges, identifying small disadvantages business (SDB) contractors, (b) establishing Government-wide priorities for strengthening strategic product and service supply-chains for such small businesses, and (c) implementing these priorities to better position federal government in building a strong breadth and depth of small business entity participation in different market segments. The first and foremost research question to be answered was developing a formal methodology on identifying priority supply chains. We describe a data-driven methodology that, (a) mines government-wide procurement databases for North American Industry Classification System (NAICS) and Product and Service Codes (PSC) for historical data related to small, disadvantaged business (SDB) participation, involving a large data set representing more than 1000 six-digit NAICS and 3000 PSC codes corresponding to SDBs, for a 12 year time-series reaching more than 20,000 at the end of the period, (b) analyzes it utilizing an ensemble of data processing, statistical, and economic analysis techniques to come up with a list of candidate supply chains. Further detailed business processes evaluation of these supply chains in the next phase of this research is intended to recommend best practices to be implemented for enhancing SDB participation in federal procurement activities.

2 - Retention Optimization in Maintenance Training Programs

Yifu Li, University of Science and Technology of China, Hefei, China, People's Republic of, Qiuwei Guo, Lindong Liu, Lifei Sheng

Training is an important business in the service sector. Although performance-enhancement training programs receive considerable attention from operations management researchers, there is a lack of guidance on designing maintenance training programs. We close this gap by investigating the modeling and optimization of a maintenance training program, considering the behavior that a customer may abandon the program if the training experience is overstressed. Customers make the participation decision of the training program based on the remembered holistic utility of the training activities. We formulate the maintenance training program design problem (MTDP), which maximizes overall service retention across all training episodes through activity selection and scheduling. By our analysis, the MTDP is a 0-1 constrained exponential sum problem (0-1 CESP), which is NP-hard. To resolve the MTDP, we introduce a novel geometric branch and bound (GB&B) algorithm that searches for the optimal solution by computing the bound of multiple dimensions over a box area. The GB&B algorithm is proven to be efficient in solving the MTDP through our computational studies, and it has the potential to resolve other 0-1 CESPs, which cannot be resolved by the branch and bound algorithm. With the optimal solution obtained by the GB&B algorithm, we investigate the properties of optimal program design. We contribute to the literature by discussing the mathematical model, solution method, and managerial insights of maintenance training program design.

3 - Strategic use of Deadlines to Counteract Present Bias

Zhixin Liu, University of Michigan-Dearborn, Dearborn, MI, United States, NICHOLAS G. HALL

We study utility loss from present bias in decision-making, resulting from procrastination and suboptimal task choice, by utilizing deadlines. Our backward recursion method for task scheduling under variable self-awareness levels yields a well-defined schedule framework. We demonstrate that task completion and utility loss respond to deadlines through self-awareness. Our model also presents a learning approach for recognizing and managing present bias. We highlight the effectiveness of deadlines, the varying impact of deadline strictness on utility, and the influence of self-awareness on self-control. Moreover, we offer practical recommendations for deadline setting and motivate managers to consider employees' present bias when allocating tasks.

4 - Online Platform's Store Brand Encroachment: the Role of Spillover Effect

Niu Yu, Wuhan Textile University, Wuhan, China, People's Republic of

In the platform economy, many online e-commerce platforms consider whether to employ the store brands to improve profits. When the national brand of the upstream manufacturer has spillover effect on the store brand, we construct a stylized model to investigate the store brand introduction choice of the online platform. By developing a framework of game theory, two common selling modes are considered: The reselling mode and the agency selling mode. Our findings indicate that when taking into account the spillover effect between the two brands, regardless of adopting a reselling or agency mode, it is always profitable for the platform to introduce the private label. Moreover, within a certain range of spillover effects, the national brand manufacturer can still benefit from the encroachment of the platform's store brand. Interestingly, we identify that in equilibrium, there exists a Pareto improvement zone wherein both the platform and the manufacturer are better off under the agency format.

5 - Uniform Parallel Machine Scheduling Game with Non-splitting Jobs

Guohua Wan, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Jin Yu

We analyze the cooperative game of the $Q_m| |\sum C_j$ problem with non-splitting-jobs for a set $N=\{1, 2, \ldots, n\}$ players, where the player i has N_i jobs. We prove the existence of the core of the game by linear program duality and characterize the core of the game. We also extend some results of Alon and Anily (2023) on the uniform parallel machine scheduling game with splitting-jobs.

MD02

Summit - 321

Manufacturing AI Case Study Competition

Award Session

Quality, Statistics and Reliability

Chair: Ran Jin, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Tom Woteki, Virginia Tech, Blacksburg, VA, United States

MD03

Summit - 322

Advanced Data Analysis for Interconnected and Distributed Systems

Invited Session

Quality, Statistics and Reliability

Chair: Elif Konyar, University of Florida, Gainesville, FL, United States

Co-Chair: Mostafa Reisi, University of Florida, Gainesville, FL, United States

1 - Multi-Physics Analysis for Additive Friction Stir Deposition (AFSD) using PINN

Zhenyu Kong, Virginia Tech, Blacksburg, VA, United States, Raghav Gnanasambandam

Metal additive manufacturing (AM) is attractive for its flexibility in designing complex parts. Additive Friction Stir Deposition (AFSD) is a solid-state AM process that deposits material by rotating a feed rod over the substrate. AFSD has several advantages over other AM techniques, including building large emissions-free parts. Modeling, monitoring, and quality control in AFSD is challenging due to the complex physics of the process. Coupled Partial Differential Equations (PDEs), particularly the heat equation and the Navier-Stokes equation, model the process to predict the thermal distribution and material flow during the process, respectively. The solutions of coupled PDEs are typically approximated with numerical simulations that consume much computational power and time. Physics-informed neural Networks (PINNs) are increasingly used to learn these solutions in a discretization-free framework. PINNs leverage the automatic differentiation of neural networks to discover the solutions of partial differential equations (PDEs). Despite their popularity, most applications do not consider multi-physics systems modeled by coupled PDEs. This work proposes coupled PINNs to learn the solutions of multi-physics systems. For AFSD modeling, the coupled PINNs have sub-networks for the heat equation and Navier Stokes, respectively. The inputs of the sub-networks are shared spatial-temporal locations, and the outputs are the corresponding thermal distribution and the material flow velocities. The loss functions of the sub-networks include the respective PDEs and corresponding boundary conditions. The proposed approach is a promising simulator for AFSD and applies to other complex systems.

2 - Selective maintenance optimization for large-scale systems with homogeneous units

Young Myoung Ko, Pohang University of Science and Technology, Pohang, Korea, Republic of, Jaehyoung Ju, Eunshin Byon

We present a novel fluid model to represent the degradation dynamics across the system, allowing for a condition-based maintenance (CBM) strategy. This strategy triggers maintenance actions when the cumulative health state of the units reaches a predetermined threshold, optimizing the long-term cost-effectiveness of maintenance operations.

Unlike typical regenerative processes where all units are restored to an 'as good as new' state, this study's selective maintenance approach repairs or replaces only those units in specific degraded states, leaving others unchanged. This non-regenerative nature presents unique analytical challenges, addressed by the proposed fluid model which effectively captures and predicts system behavior over time.

Numerical experiments validate the model's accuracy and efficiency, showing that the proposed CBM strategy outperforms traditional periodic maintenance approaches by minimizing the long-run average cost, providing significant improvements in maintenance planning for systems with many units. This approach is particularly relevant to industries like wind farms or solar parks, where many small generators require efficient maintenance management.

3 - An Adaptive Sampling Strategy for Online Monitoring of Partially Observed Networks

Yue Jiang, Purdue University, West Lafayette, IN, United States, Ana Maria Estarda Gomez

Rapid advancements in sensing and communication technologies have necessitated monitoring data streams with a network structure in various applications, including power grid maintenance, drought monitoring, and cyber-attack detection. In some applications, the network system can only provide partial information in real time due to limited data acquisition, transmission, and processing resources. In this case, it is critical to decide which node in the network to collect data at each time epoch to maximize the change detection capability. This paper proposes an adaptive sampling strategy for online monitoring of partially observed networks. First, a Gaussian process with a novel spatial kernel that exploits the global network structure is developed to capture the spatio-temporal information within the network. This information is used to guide the monitoring scheme and the adaptive sampling strategy. The main idea of the adaptive sampling strategy is to balance exploration and exploitation by designing a sampling distribution function based on the network topology to decide where to collect data at each acquisition time. The performance of the proposed framework is demonstrated through simulations and case studies.

4 - Dynamic Sensor Selections from In-Field Units for Remote Prognostics in a Cloud-Based System Center

Ying Fu, University of Wisconsin-Madison, Madison, WI, United States

The precise prognostics of operational units are crucial to avoiding catastrophic events. Units in sectors like automotive and aviation, often operating remotely, are equipped with advanced computing systems and wireless sensors for tracking, monitoring, and control. The computing systems can autonomously evaluate maintenance needs or make decisions using local sensor data, which is also transmitted to central or cloud-based systems for continuous improvement. However, high transmission costs, bandwidth constraints, and privacy concerns restrict data transmission frequency, necessitating sensor selection. Additionally, the changing operating environment requires dynamic sensor selection to enhance system robustness. In this study, we strategically select sensors to minimize prediction errors between full and partial sensor sets by utilizing the computational capabilities of the onboard system within bandwidth constraints. Due to the lack of a reliable prediction model, we propose a pilot model that starts with minimal data and dynamically updates as new data is collected on the server. Given the objective function is a complex neural network, we adopt a local linear approximation method to facilitate an effective solution. Our approach is empirically validated with a dataset from an aircraft gas turbine engine.

MD04

Summit - 323

Data Science for Cyber-Physical Security and Resilience

Invited Session

Quality, Statistics and Reliability

Chair: Dan Li, Clemson University, Clemson, SC, United States

Co-Chair: Navid Aftabi, University of Wisconsin-Madison, Madison, WI, United States

1 - A Reinforcement Learning Framework for Dynamic Watermarking in the Industrial Machine Tool Controllers

Navid Aftabi, Clemson University, Clemson, SC, United States

With worldwide growth in industrial automation and digitalization of the manufacturing environment, the manufacturing plants are no longer isolated from cyberattacks. Consequently, the industrial machine tool controllers (MTCs) are increasingly integrated with cloud applications, external sensors, and other data sources. While this transformation holds immense potential for efficiency and innovation, it simultaneously presents formidable cybersecurity challenges. The interconnectedness of MTCs with external networks makes them more vulnerable and primary attack targets for cyberattacks. This paper addresses the problem of secure MTCs by designing dynamic physical watermark signals using reinforcement learning (RL). We study the detection of replay attacks on MTCs, and introduce a watermarking RL framework for joint optimization of detection of replay attacks and control performance.

2 - Statistical Inference for Software Reliability Constrained by the Shape of the Mean Value Function

Kangan Chen, AMSS， Chinese Academy of Science, Beijing, China, People's Republic of

While parametric Software Reliability Growth Models (SRGMs) serve as a cornerstone in software reliability assessment, their reliance on known fault-detection time distributions often presents a significant limitation in practical software testing. In this study, we develop a novel shape-restricted spline estimator for quantifying software reliability, which enjoys the following advantages: 1) This estimator not only shares a key characteristic with parametric SRGMs, a feature largely underexplored in existing literature, but also obviates the need for specifying fault-detection time distributions; 2) It effectively utilizes the critical shape information of the mean value function (MVF) of fault-detection process, a detail seldom considered in prior work; 3) Unlike common non-parametric software reliability models, our shape-restricted spline estimator is capable of predicting the future behavior of failures. Furthermore, we investigate the predictive performance of the proposed methods by employing the so-called one-step look-ahead prediction method. In numerical experiment, we show that spline estimators under restriction demonstrates competitive performance compared to parametric and certain non-parametric models.

3 - Using Analytics and Streaming Data to Prioritize Vulnerability Patching and Remediation

Theodore Allen, The Ohio State University, Columbus, OH, United States, Garett Fogel, John McCarty

4 - Smart and Secure Manufacturing: Maintainable Digital Twins (DTs) via Bayesian Optimization and Active Learning

Abhishek Hanchate, Texas A&M University, College Station, TX, United States, Satish Bukkapatnam

This work proposes an implementation of Bayesian Optimization (BO), specifically an extension on Efficient Global Optimization (EGO), in an active learning context for developing maintainable Digital Twins (DTs). The motivation is driven by the challenges of effectively obtaining high-quality data for training DTs, often leading to significant deterioration in the reliability and performance of them. This becomes especially relevant for noisy environments wherein the physical system itself might change. For a given manufacturing recipe (G-code) for a machine, a residual matrix is generated by comparing the actual machine trajectory with the one followed by an initial DT in the form of an emulator. BO is then employed in order to approximate and regenerate the residual surface in least number of BO iterations, resulting in regenerated trajectories based on sampled points that follow the machine's spatial constraints. Additionally, the intermediate points between the trajectories are also employed in the next BO iteration to accelerate the convergence process, usually tracked via a criterion such as Sum of Squared Errors (SSE). The sampled points are then employed to probe the system, allowing for recursive and iterative generation of additional G-codes by stitching the points together, allowing for continuous training of the DT. Due to the noise-robust nature of EGO, the data generated are resistant to any physical system changes and are grounded in real-world machinery. The work demonstrates the novelty and potential of such frameworks for enhancing data quality and quantity for DTs.

5 - Maintenance Optimization for Multi-Station Manufacturing Systems with Quality-Reliability Chain

Kani Fu, University of Florida, GAINESVILLE, FL, United States, Minhee Kim

In many multi-station manufacturing systems, the system reliability is affected by both station reliability and product quality. Specifically, the incoming quality of the product can affect the reliability of the station. Meanwhile, the degradation of the station can cause deterioration of the product, resulting in bad quality. This interaction between product quality and station reliability is called the Quality-Reliability Chain (QR-chain). However, when studying the maintenance policy of multi-station manufacturing systems, little research has considered this dependency. Considering the characteristics of QR-chain, we formulate this problem as a mixed observability Markov decision process (MOMDP) model, in which the station's reliability is observable but the product's quality is unobservable except in the last station. By minimizing the long-run average cost from both station maintenance and product failure, we obtain the optimal maintenance policy. Several structural results are derived by the value iteration algorithm to find the optimal maintenance thresholds. A numerical study is conducted to illustrate the control insight of the optimal maintenance policy.

MD05

Summit - 324

Models for Supply Chain Resilience and Disruption

Contributed Session

Chair: Brent Moritz, Penn State University, University Park, PA, United States

1 - *Digital Technology Adoption Impact on Supply Chain Resilience*

Chunxia (Trisha) Zhang, Kennesaw State University, Marietta, GA, United States

Supply chain resilience has garnered significant attention from scholars and practitioners, particularly in the face of global environmental disasters and pandemics like COVID-19. This study delves into the drivers behind firms' technology adoption to fortify their supply chain resilience, the hurdles faced during this adoption process amidst complexity and uncertainty, and the integrated role of technology, human capital, and organizational structure in augmenting resilience.

Drawing on the resource-based view theory, my research investigates the intricate relationship between adopting digital supply chain technologies, the capabilities of supply chain integration, and the resulting impact on supply chain resilience. Through in-depth interviews with top-level supply chain executives, I identify four key constructs shaping firms' resilience strategies:

Fit: The alignment between customer demands and technological competencies influences technology adoption decisions, subsequently impacting resilience strategies.

People – Labor: Labor constraints significantly influence technology adoption decisions, thus shaping resilience strategies.

Technology Competencies: Proficiency in digital supply chain technologies correlates with organizational structure and workforce capabilities, affecting resilience.

Supply Chain Integration: Integrating traditional and emerging technologies internally and externally contributes to customer satisfaction, consequently influencing technology adoption and resilience strategies.

My study contributes theoretically and practically by offering insights into the complex interplay between technology adoption, supply chain integration, and resilience. It provides actionable guidance for managers in navigating technology adoption decisions to enhance supply chain resilience. Additionally, while acknowledging the study's limitations, such as sample scope, I suggest avenues for future research, including exploring the impact of contextual factors on the technology adoption-resilience relationship.

2 - Sustainable Supplier Selection Decisions: Choices and Tradeoffs of Practicing Managers

Brent Moritz, Penn State University, University Park, PA, United States, Ben Fahimnia, Andrew Collins

Selecting environmentally sustainable suppliers is a vital and visible supply chain function. However, suppliers differ in cost and performance across different dimensions of sustainability. In addition, suppliers are selected by managers with heterogenous preferences. Little is known about how these factors interact. Using a large sample of experienced supply chain managers, we conduct a discrete choice supplier selectin experiment to investigate how practitioners value five key supplier attributes: *cost, waste management, carbon efficiency, green product and process design, and organizational support for environmental sustainability*. We assess the willingness to pay (WTP) for different dimensions of sustainability. Our results show that experienced practitioners were willing to pay the most for supplier improvements to waste management and carbon efficiency, and somewhat less for green product and process design, and organizational support for sustainability. We found modest differences by organizational level, education, and gender, and find that individuals with sustainability experience were willing to pay substantially more for higher environmental performance. Surprisingly, the WTP is highly non-linear: Managers with the highest individual environmental preferences were willing to pay nearly ten times more for improvements to sustainability. We show which key dimensions of sustainability are favored by practitioners, and that they are favored more (or much more) by individuals with high environmental performance orientation.

3 - Lean Management in Supplier Selection Process

Edward Lee, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Edward Lee

Lean Management in Supplier Selection Process

Supplier selection has played a significant role in supply chain management. A trustful and reliable supplier is essential to the success of a business. An effective and efficient supplier selection process is therefore critical. To provide the products to satisfy the customers, the quicker to conclude a supplier the better. Selecting and evaluating qualified suppliers is complicated. The longer the selection process, the more logistics cost will be consumed. The lean management approach can help companies save costs in the selection process, such as using the lowest cost to evaluate the availability and adaptability of potential suppliers and reducing costs in investigating competitiveness and capability. This paper aims at introducing how lean management techniques can be applied to the supplier selection process. The lean tools discussed in the paper not only result in cost savings, but effects of long-term supplier relationships. New supplier selection strategies will also be considered within the value stream to ensure excellence in selection procedures, in terms of improvement within quality, cost, delivery, and design processes.

MD06

Summit - 325

Emerging Markets: Challenges and Opportunities

Contributed Session

Chair: N ravichandran, IIM Ahmedabad, C 501 casa Vyoma Vastrapur, Ahmedabad, 380015, India

1 - Balancing Productivity Growth and Environmental Efficiency Goals in Indian Manufacturing: Some Alternative Scenarios

Kankana Mukherjee, Babson College, Wellesley, MA, United States

Soon after Prime Minister Narendra Modi came into power in 2014, the ‘Make in India’ initiative was launched by the Government of India which aimed at increasing the share of manufacturing in the Indian economy to help propel economic growth and create millions of jobs. But an increasing manufacturing production places very large demand on energy. As most industries in India still use traditional sources of energy, they are big contributors to overall greenhouse gas emissions. The goals of increasing manufacturing production and employment while at the same time reducing energy use conflict with each other. This study investigates these tradeoffs by examining data for the major energy intensive industries across the states in India over the period 2013-14 to 2019-20. Using data from Annual Survey of Industries, Government of India as well as other sources to create industry specific data and utilizing Data Envelopment Analysis and directional distance functions, I propose several alternative models for reconciling the conflicting goals in Indian manufacturing. Results show that significant potential for simultaneous increase in output and reduction in energy use exists in several states. The results from the empirical analysis should be helpful in providing policy directions for the government as well as strategic insights for the firms in the energy intensive industries in India.

2 - Does Corporate Governance Improve the ESG performance? Evidence From an Emerging Market

Surya Bhushan Kumar, Indian Institute of Management Indore, Indore, India

Corporate firms disclose their ability and performance through various mandatory and non-mandatory reports. While Annual reports are mandatory, there is a rising trend of disclosing more with some non-traditional reports. ESG report is one such non-mandatory report that provides enriched information to all firm stakeholders. The usability of such ESG reports is more important for firms in emerging markets due to constrained environment for information disclosure and other regulatory challenges. It also aggravates the problem of information asymmetry for investors in their decision-making. Here it will be interesting to see the role of improved corporate governance. In this line, I am exploring in this study the role of corporate governance as an antecedent for ESG performance in emerging markets. Academic literature also suggests that components of ESG vary in their information content and importance. Additionally, it explores the specific element that contributes more to ESG performance. The findings will be helpful to all stakeholders for better decision-making. As mandatory ESG performance in emerging markets is relatively rare compared to developed markets, the finding of this study could enrich the extant related literature.

3 - Digitisation and its impact on the largest employment insurance scheme in the world

Ravichandran Narasimhan, Indian Institute of Management-Ahmadabad, Ahmedabad, India, Balaraman Rajan

The largest employment insurance in the world operating in India has enhanced consumer focus by digitisation of their process. This inhouse effort consisted of Business process reengineering, structural changes, automation and structural changes in the operations. The change management involved working with employees of the organisation and other stake holders like members and member employers .The impact of the digitisation was ease of doing business by members and rapid response time. This paper documents this experience as a unique case study in under managed sectors.

MD07

Summit - 327

Data Science in Healthcare and Policy

Flash Session

Chair: Alireza Boloori, University of Washington, Tacoma, WA, United States

1 - Comparative Analysis of Comorbidity Networks Before and During the COVID-19 Pandemic

nguyen phan, university of Arkansas, fayetteville, AR, United States, Shengfan Zhang

The COVID-19 pandemic has caused unprecedented impacts on a global scale, including public health system exhaustion, supply chain disruption, economic downturn, service industry decline, lasting behavioral changes, and many more. The pandemic can cause a significant and rapid impact because of the lack of preparedness and tools to detect abnormal changes. To develop such tools, we propose a comorbidity network analysis (CNA) approach, using the National (Nationwide) Inpatient Sample (NIS) to map the interconnections between various comorbidities within patient cohorts. In our analysis, we include diverse demographic and socioeconomic factors, giving us a detailed view of the pandemic’s disparate effects on different populations. To examine how the pandemic has changed the complex comorbidity network dynamics, we constructed and compared different comorbidity networks for each quarter of the year 2020 and pre-pandemic years 2018-2019. Through our observation study, we seek to observe any correlations and trends among concurrent diseases, providing valuable insights into how the pandemic has changed the complex comorbidity network dynamics. Understanding the nature of the virus and how the pandemic‘s effect on different populations can contribute to practical healthcare applications and strategic responses to pandemics.

2 - AI-Powered Tool For Early Prevention Of Type 2 Diabetes In Adolescents

Mahima Naznin, University Of Missouri- Columbia, Columbia, MO, United States, Sharan Srinivas

Early interventions during adolescence and young adulthood are crucial to averting over 50% of Type 2 Diabetes (T2D) cases, which are expected to nearly triple in the US by 2060. Unfortunately, there is currently no screening tool available to predict adolescents' future T2D risk. This research aims to develop a machine learning-based screening tool for predicting short-term and long-term T2D risk among adolescents by accounting for their lifestyle choices and socioecological, physiological, and psychological factors. We evaluated and compared the area under the receiver operating characteristic curve (AUC-ROC) for several state-of-the-art machine learning algorithms. Results indicate XGBoost to yield the best performance with an AUC-ROC of 0.72, suggesting good discriminating capability in predicting high-risk cases up to 15 years in advance.

3 - Identifying Genetic Variants for Obesity Incorporating Prior Insights: Quantile Regression with Insight Fusion for Ultra-high Dimensional Data

Jiantong Wang, University of Cincinnati, Cincinnati, OH, United States, Heng Lian, Yan Yu, Heping Zhang

Obesity is widely recognized as a critical and pervasive health concern. We strive to identify important genetic risk factors from hundreds of thousands of single nucleotide polymorphisms (SNP) for obesity. We propose and apply a novel Quantile Regression with Insight Fusion (QRIF) approach that can integrate insights from established studies or domain knowledge to simultaneously select variables and modeling for ultra-high dimensional genetic data, focusing on high conditional quantiles of body mass index (BMI) that are of most interest. We discover interesting new SNPs and shed new light on a comprehensive view of the underlying genetic risk factors for different levels of BMI. This may potentially pave the way for more precise and targeted treatment strategies. The QRIF approach intends to balance the trade-off between the prior insights and the observed data while being robust to potential false information. We further establish the desirable asymptotic properties under the challenging non-differentiable check loss functions via Huber loss approximation and nonconvex SCAD penalty via local linear approximation. Finally, we develop an efficient algorithm for the QRIF approach. Our simulation studies further demonstrate its effectiveness.

4 - Analysis of Social Determinants of Health to Risk-Stratify Type II Diabetes Pathways

Yukti Kathuria, University of Massachusetts Amherst, Amherst, MA, United States, Muge Capan, Kristen Miller, Elizabeth Selden, William Gallagher

Type II diabetes is a significant public health burden with high prevalence and high cost. Non-clinical factors, such as demographics (e.g., race, ethnicity, gender) and social determinants of health (SDOH) (e.g., insurance, marital status, health literacy) have a strong influence on individuals' diagnostic trajectories in primary care. However, most decision analytical models focus on clinical factors to determine the risk of missed opportunities and undesired health outcomes. This study focuses on quantifying and categorizing patient subpopulations within primary care settings to study the impact of these SDOH factors and demographics on patient outcomes. Descriptive and statistical analyses were applied to de-identified retrospective longitudinal Electronic Health Record (EHR) data (e.g., demographics, labs, vital signs, diagnoses, and health care utilization) of adult patients in selected primary care practices over a 6-year study period (01/2017-12/2023) derived from a healthcare delivery system in the mid-Atlantic, serving a racially, ethnically, and socioeconomically diverse patient population residing in urban, rural, and suburban settings. By employing clustering techniques, we identified distinct subpopulations and their trajectories based on their SDOH characteristics and Type II diabetes pathways. This approach not only enhances our understanding of how social factors influence each other and diabetes diagnosis pathways, but also aids in risk prioritization in primary care diagnostic decision processes. Our preliminary findings suggest significant variance in risk of adverse diagnostic trajectories among subpopulations. These insights will help providers to identify at-risk patients more effectively, thereby optimizing resource allocation and improving primary care diagnostic decision making.

5 - Benchmarking Fairness of Genetic Risk Score Models for Early-stage Prediction of Type 1 Diabetes from the TEDDY study

Feng Lin, University of Washington, Seattle, WA, United States, Congjing Zhang, Kendra Vehik, Hemang Parikh, Mingqian Li, Richard Oram, Xiaoning Qian, Shuai Huang

Type 1 diabetes (T1D) is one of the most prevalent and serious chronic immune-mediated diseases in childhood which has a great impact on the patients' life expectancy and their quality of life. The rapid advances in machine learning (ML) have fostered the use of various ML models for building prediction genetic risk score (GRS) models for the onset of T1D in an early stage. However, the existence of ethnoracial inequity has cast a shadow over the use of these GRS models in practice. Thus, it is important to comprehensively evaluate the potential unfairness in these algorithms across different ethnoracial groups. By performing a secondary analysis of the The Environmental Determinants of Diabetes in the Young (TEDDY) study data with the state-of-the-art ML models, we show that current ML GRS models demonstrated considerable disparities across ethnic groups. Our findings highlight the importance of considering fairness in the development of ML GRS models for T1D prediction as well as in future study design of T1D research.

6 - An interpretable and deterministic model to improve legal decision-making in criminal justice reform

Aparna Komarla, Redo.io, Berkeley, CA, United States

In 2018, California became the first state to pass a Prosecutor-Initiated Resentencing (PIR) law, which grants prosecutor offices the power to revisit multi-decade and life sentence cases. But given that there are over 100,000 people in California's prisons, lawyers and county officials need efficient and accurate ways to identify eligible candidates for PIR instead of manually reviewing piles of paperwork. Redo.io (github.com/redoio) is solving this problem with an open-source eligibility model to filter cases and apply legal expertise efficiently. We designed a rules-based and deterministic model as opposed to a black-box predictive model that embodies the principles of fairness, accountability and transparency. Legal experts help design and fine-tune eligibility rules in an iterative review process. Our model therefore augments legal knowledge about the applicability of PIR as opposed to replacing such expertise. Redo.io has deployed the eligibility model for the Three Strikes Project at the Stanford School of Law (https://law.stanford.edu/three-strikes-project/), a legal clinical seminar where law students represent individuals serving life sentences for nonviolent crimes under California's Three Strikes Law. By supplementing a lawyer's decision to examine a case for resentencing instead of automating the end-to-end process, our model helps stakeholders make high-impact decisions efficiently while preventing existing biases in our society from being exacerbated. It therefore exemplifies a framework for meaningful human-computer interaction in high-stakes and high-risk decision making processes.

7 - Early Detection of Problematic Cannabis Use Among Young Adult Users in Los Angeles Using a Recurrent Neural Network Approach

Jiyang He, University of Southern California, Los Angeles, CA, United States, Shinyi Wu, Carolyn Wong

The increasing prevalence of cannabis use worldwide underscores the need for a thorough understanding of cannabis dependence, highlighting the importance of early detection to mitigate long-term health consequences. This study aims to detect problematic cannabis use (PCU) in the dynamic environment of Los Angeles County, one of the nation's largest markets for liberalized cannabis use. These findings could assist healthcare professionals in cannabis intervention programs by identifying risk factors and warning signs of PCU, thus facilitating the implementation of early intervention strategies.

This research employs a Recurrent Neural Network (RNN) model using data from a longitudinal study of cannabis-using young adults aged 18-26 from 2014 to 2023, which includes eight waves of data collection. The dataset incorporates sociodemographic data, self-reported substance use, and psychosocial assessments. The determination of cannabis dependence among participants utilizes a diagnostic criterion based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).

Our approach integrates these multi-dimensional data sources with a deep learning architecture to model the potential onset of PCU. The RNN model processes time-variant features through multiple layers to capture dynamic changes over time. Additionally, we use a Lassonet-RNN model to reduce input feature dimensions and determine the importance rankings of characteristic variables. We further evaluate the effectiveness of the RNN model by comparing it with traditional machine learning methods.

These analyses provide insights to inform preventive health strategies and practical decision-making in clinical settings, highlighting the potential of deep learning models and the value of utilizing a longitudinal data structure.

8 - Analyzing the Impact of Cannabis Legalization on Health, Societal, and Workplace Outcomes

Alireza Boloori, University of Washington, Tacoma, WA, United States, Iman Attari, Michael Freeman, Rupinder Jindal

The recreational use of cannabis has been legalized in several U.S. states over the last decade. Utilizing patients' claims data, we investigate the impact of these legalizations across states on various healthcare (e.g., hospital resources utilization), societal (e.g., domestic violence), and workplace (e.g., injuries/disabilities) outcomes. We employ a staggered Difference-in-Differences method to test whether or not such legalizations have impacted the foregoing outcomes (and, if so, under what circumstances). Leveraging findings from our numerical experiments, we provide various insights for Federal/state-level healthcare policy makers that could help them in approaching cannabis-related legislation.

MD08

Summit - 328

MAS Student Interactive Presentations

Invited Session

Military and Security

Chair: Isabella Sanders, Dept. of Systems Engineering, United States Military Academy, Newburgh, United States

1 - Risk Optimization in Food Supply Chains

Ryan Jackovic, Purdue University, West Lafayette, IN, United States

Current methodologies in supply chain risk assessment often overlook the interconnectedness and cumulative impacts of risks across the supply chain. This paper proposes a novel methodology that addresses these gaps by integrating supplier selection and order allocation into a comprehensive supply chain risk assessment framework. The proposed model evaluates potential suppliers based on their production capacities, costs, and risks, considering factors such as location risks, transportation vulnerabilities, and financial stability. The technique identifies optimal paths that balance resilience and efficiency using an optimization model that minimizes shutdown costs, purchasing expenses, and carbon emissions. Traditional multi-criteria approaches may complicate decision-making with multiple trade-offs, while our model simplifies the process by providing a single, interpretable solution. This approach is demonstrated through a case study on food supply chains.

2 - Stochastic Military Port Selection

Lennon Conson, Air Force Institute of Technology, Wright-Patterson AFB, OH, United States, Brian Lunday, Matthew Robbins

The military seaport selection problem involves determining the seaports of embarkation (SPOEs) for the transshipment of military cargoes from home stations to theaters of operation to minimize cumulative travel times. Whereas conventional, commercial seaport selection for transshipment has primarily focused on reducing economic costs, this research examines a military-focused, temporal-domain optimization of seaport selection for military cargo. This problem has several complicating factors of interest: shipments originating at respective home stations traverse either rail or road networks to reach an SPOE; transporting ships must be managed to carry shipments from SPOEs to seaports of debarkation; and the identification of seaports are first-stage decisions, whereas shipment routing and timing entail second-stage decisions subject to parametric uncertainties. This study examines model tractability for a representative scenario, identifying robust, resource-efficient strategic deployment decisions.

3 - Supplier Selection in DoD Supply Chains

D'Andre Tobias, Georgia Institute of Technology, Atlanta, GA, United States

Plant closures have always been a point of concern for supply chain purchasers within the Department of Defense (DoD). At any time, a supplier plant could close due to factors such as natural disasters, strikes, pandemics or financial distress. In times of crisis, the risks of such factors can increase exponentially. When identifying annual contracts, it is important for supply chain purchasers to assess the risk of potential suppliers. In this paper, we apply a supplier risk assessment model proven in literature to a case study within the DoD during a time of crisis – the COVID-19 pandemic. In this case study, we use Cox Proportional Hazards Model to analyze the potential factors that influence a plant closure for at-risk suppliers determined by the supplier risk assessment. Our results demonstrate that applying the cox-proportional hazards model can aid the DoD by reducing plant closures and saving taxpayer-dollars.

MD09

Summit - 329

Autonomous Vehicle Operations

Invited Session

Auctions and Market Design

Chair: Daniel Freund, MIT, Cambridge, MA, United States

Co-Chair: Ilan Lobel, New York University, New York, NY, United States

1 - Cooperation for Scalable Supervision of Autonomy in Mixed Traffic

Cathy Wu, Massachusetts Institue of Technology, Cambridge, MA, United States, Cameron Hickert, Sirui Li

Advances in autonomy offer the potential for dramatic positive outcomes in a number of domains, yet enabling their safe deployment remains an open problem. This work's motivating question is: In safety-critical settings, can we avoid the need to have one human supervise one machine at all times? The work formalizes this scalable supervision problem by considering remotely located human supervisors and investigating how autonomous agents can cooperate to achieve safety. This article focuses on the safety-critical context of autonomous vehicles (AVs) merging into traffic consisting of a mixture of AVs and human drivers. The analysis establishes high reliability upper bounds on human supervision requirements. It further shows that AV cooperation can improve supervision reliability by orders of magnitude and counterintuitively requires fewer supervisors (per AV) as more AVs are adopted. These analytical results leverage queuing-theoretic analysis, order statistics, and a conservative, reachability-based approach. A key takeaway is the potential value of cooperation in enabling the deployment of autonomy at scale. While this work focuses on AVs, the scalable supervision framework may be of independent interest to a broader array of autonomous control challenges. This work is published in the Transactions on Robotics in 2023.

2 - Supply and Capacity Planning For Autonomous Delivery Vehicles

Neda Mirzaeian, The University of Texas at Dallas, Richardson, TX, United States, Milind Dawande, Soraya (Nadia) Fatehi, Ganesh Janakiraman

We investigate optimal staffing and pricing strategies for a food delivery platform incorporating both human drivers and autonomous delivery vehicles (ADVs) under a maximum delivery time guarantee. Modeling the platform's objective of minimizing its operational costs as an optimization problem, we show nuanced interplays between drivers' wages, vehicle wear and tear, and ADV leasing costs. A comparison is then drawn between the feasibility and optimal number of servers in this case and a benchmark where the platform solely relies on human drivers. We further analyze the potential of ADVs to entirely replace human drivers in the food delivery industry. Our findings suggest that optimal resource allocation may require the simultaneous utilization of both human drivers and ADVs, even in the presence of an abundant supply of ADVs. Our study highlights scenarios where human drivers remain indispensable, challenging assumptions about the future dominance of ADVs. In addition, we explore strategic staffing and pricing for non-peak and peak demand periods. Our results challenge conventional beliefs regarding resource utilization during non-peak and peak periods, demonstrating the advantages of employing human drivers even during non-peak hours. Our findings offer valuable insights for food delivery platforms, ADV suppliers, and human drivers, assisting in optimal resource deployment and providing foresight into the evolving landscape of food delivery services.

3 - On the Supply of Autonomous Vehicles in Platforms

Ilan Lobel, New York University, New York, NY, United States, Daniel Freund, Jiayu (Kamessi) Zhao

The likely large-scale deployment of autonomous vehicle (AV) technology in the near future has the potential to fundamentally change the transportation landscape. Due to the high cost of AV hardware, the most likely path to widespread AV use is via platforms that can sustain high utilization, such as ride-hailing and delivery services. In this paper, we consider four potential operational models to commercialize AVs, which we model as a supply chain game between a platform, an AV supplier, and human drivers that join as individual contractors (ICs). We find that apart from an AV-only platform, all deployment models are subject to a risk of AV underutilization due to the need to maintain the ICs' utilization sufficiently high to ensure ICs remain engaged. In non-integrated models, this risk can propagate backwards in the supply chain and thereby negatively affect its efficiency. For the open platform the efficiency loss compared to an integrated platform can be unbounded. We then study how usage commitments between the platform and the supplier can overcome the unbounded efficiency loss. Though the uncontracted open platform may perform worse than the other models, an open platform equipped with a usage contract offers substantial efficiency benefits compared to the AV-only/AV leasing platforms.

4 - Getting out of Your Own Way: Introducing Autonomous Vehicles on a Ride-Hailing Platform

Andrew Frazelle, The University of Texas at Dallas, Richardson, TX, United States, Francisco Castro

We consider a ride-hailing platform's strategy to recruit human drivers while also operating a private fleet of AVs. The platform sets the human-driver wage and its AV fleet size, and human drivers make strategic joining decisions based on a rational anticipation of their expected earnings. We show that having the option to augment its AV fleet after observing human participation levels can, counterintuitively, hurt the platform's bottom line. Human drivers anticipate that the platform may reactively acquire more AVs; this reduces their expected earnings, so the platform must offer higher wages to attract a given number of drivers. The higher wage implies that the platform should acquire even more AVs, leading to a feedback loop or "race to the top" of increasing wages and increasing AV acquisition. The race to the top hurts the platform's profits by effectively preventing it from attracting more than a limited number of human drivers and increasing the cost of attracting a given number. Our findings reveal the importance of the platform's credibility with its drivers. The ride-hailing platform can succeed in avoiding the race to the top only to the extent that it can convince human drivers that it will not reactively acquire a large number of AVs. Otherwise, significant profit losses await.

MD10

Summit - 330

Learning and Markets

Invited Session

Auctions and Market Design

Chair: Negin Golrezaei, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Qinyi Chen, MIT, Cambridge, MA, United States

1 - Nearly Minimax Optimal Regret for Multinomial Logistic Bandit

Min-hwan Oh, Seoul National University, Seoul, Korea, Republic of, Joongkyu Lee

In this paper, we study the contextual multinomial logit (MNL) bandit problem in which a learning agent sequentially selects an assortment based on contextual information, and user feedback follows an MNL choice model. There has been a significant discrepancy between lower and upper regret bounds, particularly regarding the maximum assortment size K. Additionally, the variation in reward structures between these bounds complicates the quest for optimality. Under uniform rewards, where all items have the same expected reward, we establish a regret lower bound of $\Omega(d\sqrt{T/K})$ and propose a constant-time algorithm, OFU-MNL+, that achieves a matching upper bound of $\tilde{O}(d\sqrt{T/K})$. Under non-uniform rewards, we prove a lower bound of $\Omega(d\sqrt{T})$ and an upper bound of $\tilde{O}(d\sqrt{T})$, also achievable by OFU-MNL+. Our empirical studies support these theoretical findings. To the best of our knowledge, this is the first work in the contextual MNL bandit literature to prove minimax optimality --- for either uniform or non-uniform reward setting --- and to propose a computationally efficient algorithm that achieves this optimality up to logarithmic factors.

2 - Online Learning for Increasing Parental Engagement On Ed Tech Platforms

Divya Singhvi, New York University, New York, NY, United States, Alex Akira Okuno, Somya Singhvi

Ensuring universal access to quality education, particularly in developing countries, remains a formidable challenge. A key issue that affects educational outcomes among low-income households is limited parent involvement in their kids' education. However, EdTech platforms can help develop cheap and effective solutions to resolve this problem. In this work, we discuss our collaboration with Rocket Learning, one of India's leading non-profit EdTech organization that collaborates with various states- and the central-government of the country to create and manage digital communities (WhatsApp Groups) to deliver educational content to parents. In collaboration with Rocket Learning, we run a series of experiments to design, implement and optimize nudges to increase parental engagement on these digital groups using various data-driven online and offline learning methods. Our findings suggest that carefully delivered nudges can substanitally improve parental engagement. Furthermore, we also present practical insights from implementing data-driven methods on the field.

3 - Learning in Multi-Unit Pay-as-Bid Auctions

Rigel Galgana, Massachusetts Institute of Technology, Cambridge, MA, United States, Negin Golrezaei

Motivated by Carbon Emissions Trading Schemes, Treasury Auctions, Procurement Auctions, and Wholesale Electricity Markets, which all involve the auctioning of multiple homogeneous units, we consider the problem of learning how to bid in repeated multi-unit pay-as-bid auctions. In each of these auctions, a large number of(identical) items are allocated to the largest submitted bids, where each agent's payment is the sum of their winning bids.

The problem is challenging due to the combinatorial nature of the action space. We first focus on the simpler offline setting, where the bidder optimizes their vector of bids with respect to the past submitted bids by other bidders. We solve this via a polynomial time dynamic program (DP) that essentially decouples the utility across units. We leverage the DP structure to design online learning algorithms with polynomial complexity and regret. We accompany these results with a regret lower bound, which matches the linear dependency in the number of units.

Our numerical simulations show that the market dynamics generally converge to a welfare maximizing equilibrium where bidders submit uniform bids when bidders behave according to our proposed learning algorithms. Additionally, we find that the pay-as-bid auction consistently generates significantly higher revenue compared to the more popular uniform price auction. This advantage positions the pay-as-bid auction as an appealing auction format in settings where earning high revenue holds significant social value, such as in Carbon Emissions Trading Schemes.

4 - Stochastic Multi-armed Bandits with Strongly Reward-Dependent Delays

Yifu Tang, University of California, Berkeley, Berkeley, CA, United States, Yingfei Wang, Zeyu Zheng

There has been increasing interest in applying multi-armed bandits to adaptive designs in patient-centric clinical trials. Existing literature either assumes no delay in observing a patient’s treatment reward, or considers delays that are independent or weakly correlated with the reward. However, certain applications give rise to situations where the delay is strongly correlated with the reward, or even that the delay is exactly the same as the reward (e.g., survival response). We formulate a new set of stochastic multi-armed bandit (MAB) problem with strongly reward-dependent delays, and provide algorithm design and associated theoretical guarantees in terms of upper and lower bounds. We show the central limit theorems with respect to the rewards for post-experiment inference.

5 - Achieving $\tilde{O}(1/\epsilon)$ Sample Complexity for Constrained Markov Decision Process

J Jiang, Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Yinyu Ye

We consider the reinforcement learning problem for the constrained Markov decision process (CMDP), which plays a central role in satisfying safety or resource constraints in sequential learning and decision-making. In this problem, we are given finite resources and a MDP with unknown transition probabilities. At each stage, we take an action, collecting a reward and consuming some resources, all assumed to be unknown and need to be learned over time. In this work, we take the first step towards deriving optimal problem-dependent guarantees for the CMDP problems. We derive a logarithmic regret bound, which translates into a O(κϵ·log^2(1/ϵ)) sample complexity bound, with κ being a problem-dependent parameter, yet independent of ϵ. Our sample complexity bound improves upon the state-of-art O(1/ϵ^2) sample complexity for CMDP problems established in the previous literature, in terms of the dependency on ϵ. To achieve this advance, we develop a new framework for analyzing CMDP problems. To be specific, our algorithm operates in the primal space and we resolve the primal LP for the CMDP problem at each period in an online manner, with adaptive remaining resource capacities. The key elements of our algorithm are: i). an eliminating procedure that characterizes one optimal basis of the primal LP, and; ii) a resolving procedure that is adaptive to the remaining resources and sticks to the characterized optimal basis.

MD11

Summit - 331

Digital Platform Applications

Contributed Session

Chair: Ajit Pratap Singh, IIM Rd, Prabandh Nagar,, 226013

1 - Resolving Copyright Infringement of Derivative Videos

Ying Wang, Northeastern University, Liaoning, China, People's Republic of, Jennifer shang

Short video platforms have witnessed a remarkable growth in recent years. Derivative video, a specific type of short video created from longer-form content, appeals to a segment of the long video audience but also poses a risk of copyright infringement. In a stylized model we investigate whether a long video platform should collaborate with a short video platform when facing potential competition from derivative short videos. We evaluate two scenarios: collaboration, where the short video platform compensates for long video copyrights; and competition, where the two parties compete and the long video platform actively detects and addresses infringement on the short platform. Our research generates valuable managerial insights for the video industry, empowering decision-makers to make informed strategic choices.

2 - A Game Theoretic Analysis of Content Provision Strategies for Video Streaming Services

Kengo Kakimizu, Keio University, Yokohama, Japan, Nobuo Matsubayashi

In recent years, competition in the video streaming service industry has intensified. As a result, Netflix, which initially only offered ad-free content, has recently launched a new ad-supported plan as part of its competitive strategy. Motivated by such real-world examples, this study conducts game-theoretic analysis to investigate under which conditions adding an ad-supported plan to an ad-free plan is profitable in a duopolistic market where the rival firm also offers its own ad-supported plan. We assume that consumers are heterogeneous in their sensitivity to advertising, which implies that plans offered at different levels of ads are vertically differentiated. We find that unlike in a standard vertical differentiation model for information goods, where a product line is never optimal due to severe cannibalization effects, the addition of an ad-supported plan can be sustained in equilibrium. Specifically, the firm benefits from adding this plan when the content quality of the ad-supported plan is moderately low and the intensity of the network effect on advertisers is high relative to that of the rival's plan.

3 - Privacy, Data and Competition: The Case of Apps For Young Children

Grazia Cecere, Institut Mines Telecom , Evry Courcouronnes, France, Catherine Tucker, Vincent Lefrere

This study examines the impact of firm size on the privacy protections offered to consumers, particularly through apps targeted at young children—a group for whom privacy concerns are paramount. Using weekly data from the Google Play Store over a three-year period, we explore whether larger firms exploit their scale to amass more data or if smaller firms exhibit less rigorous data protection practices. Our findings indicate that larger app developers implement more robust privacy protections, requesting less sensitive data from users compared to smaller developers. Large developers originated from developing countries with low privacy regime show a tendency to adopt similar levels of data collection. This pattern suggests a trend towards a universal standardization of privacy practices among large developers, who maintain consistent privacy standards across various regulatory environments.

4 - Optimizing Profitability in Platform Ecosystems: Integrating Pricing, Differentiation, and Cross-Network Effects

Ajit Pratap Singh, Indian Institute of Management, Lucknow, Lucknow, India, Suresh Jakhar

Our research delves into the optimal pricing strategy within a duopolistic media platform market, considering the interplay of cross-network effects among viewers, content creators, and advertisers. Employing a game-theoretical linear city model, we analyze the dynamics among these interconnected stakeholders across three sides. Our investigation examines how platform pricing adjusts with the presence of multi-homing on one or more market sides. The insights obtained from our study underscore strategic opportunities for competitive platforms to leverage pricing, differentiation, and cross-network effects effectively to optimize their profits. Further, we evaluate the relevance of existing literature findings within our context, identifying instances of both alignment and deviation across various scenarios. Our findings suggest that under specific conditions, multi-homing can yield benefits for both market agents and platforms. We explore scenarios wherein multiple sides of the market engage in multi-homing concurrently, proposing that platforms should incentivize one side to multi-home only when other sides also participate in multi-homing. Furthermore, we outline criteria for offering free access or implementing subscription fees for market agents, offering practical guidance for platforms operating within a competitive landscape in the presence of cross-network effects and multi-homing.

MD12

Summit - 332

Pricing Strategies and Consumer Insights

Flash Session

Chair: Heddy Barale

1 - Flexibility in an Integrated Green Hydrogen/Ammonia Production Facility to Interface with Volatile Electricity Prices

Heddy Barale, MIT, Cambridge, MA, United States

This study explores the integration of nuclear power plants (NPPs) with thermal energy storage (TES) systems on an integrated green hydrogen/ammonia production system to facilitate participation in multi-commodity markets, thereby improving economic competitiveness. The research focuses on two types of NPPs: a light-water reactor (LWR) and a Molten Salt Reactor (MSR). These were analyzed to demonstrate the economic viability of thermally balanced energy storage coupling designs for thermal power extraction. The coupling of each reactor with a TES system was evaluated within the Electric Reliability Council of Texas (ERCOT) market framework.

A stochastic optimization approach was employed, utilizing an autoregressive moving average (ARMA) model to interpret price signals from the ERCOT market. Advanced simulation tools, including the Risk Analysis Virtual Environment (RAVEN) and its dispatch optimization plugin, the Holistic Energy Resource Optimization Network (HERON), were used for dispatch and capacity optimization. These tools leveraged data from the ARMA model to optimize system design and operations.

The findings will guide the design and characterization of an integrated system, assessing the potential of flexible operation in selling electricity. Furthermore, the feedback from market participation will influence the sizing of the reactor, thermal storage, and power block. Future work will extend to optimizing thermal storage in conjunction with water, hydrogen, and ammonia production, as well as managing the storage of these intermediate products.

2 - Sock Matching with Loss

James Crumpacker, University of Illinois, Champaign, IL, United States, Zachary Hornberger, Sheldon Jacobson

The classic Sock Matching Problem states: consider a set N whose elements (e.g., socks) are partitioned into two distinct but equal sized complementary classes (e.g., left and right socks), where the elements within one class have a unique corresponding item in the other class. Elements are selected from N sequentially at random and held (e.g., on a laundry table). Once a matching pair has been obtained, the pair is removed from the table. The problem’s traditional framing seeks the expected number of unmatched elements held over the course of the drawing process. To extend the applicability of the sock matching problem, this research considers a generalization where items in N are subject to random loss and determines exactly the expected number of draws until the first match is obtained. We present a formulation of the classic sock matching problem as a Markovian process, which offers additional insights relative to existing approaches. The model is then extended to account for random loss and calculate the expected number of draws until a match is obtained. This problem models the process of collectors seeking to acquire matching sets of collectible items such as rare coins and banknotes; this problem may be considered as a variation of the coupon collector problem.

3 - Determining the Optimal Range of a Revenue-sharing Ratio for Multiperiod Consignment Contracts under Retailer Competition

SEONG-HYUN NAM, University of North Dakota, Grand Forks, ND, United States

A Consignment Contract with Revenue Sharing model has been one of the strategic business methods used worldwide for decades as it provides manufacturers with a shared risk/reward-based revenue-generating approach to selling products through intermediary retailers. Our study aims to develop a quantitative analysis model in which the manufacturer designs a multi-period consignment contract that includes retailer pricing and volume decisions. Eligible retailers can participate in this consignment arrangement by proposing a revenue share ratio contract. However, retailers must compete with each other to reach an agreement with the manufacturer, and negotiations between the manufacturer and retailers are necessary to establish acceptable terms. Under these circumstances, our study focuses on the following questions: How does competition among retailers affect the upper bound of the range of revenue-sharing rates in multi-period consignment arrangements? Can sharing demand information between the manufacturer and selected retailers increase supply chain efficiency by improving the manufacturer's forecasting accuracy? If so, how can this lead to adjustments in the revenue share range for the benefit of all parties in recurring contractual relationships? How can sales volatility caused by demand uncertainty in multi-period contracts affect the financial risks associated with future profit and revenue sharing between the parties? This study also specifically examines a scenario where increased market uncertainty leads to panic buying or overstocking, as observed during the COVID-19 pandemic.

4 - Unveiling the Dynamics of Pricing Strategies and Framing Feedback Mechanisms on Consumer Satisfaction: A Laboratory Experiment

Kyaw Zay Ya, National Cheng Kung University, Tainan, Taiwan, Wei-Shiun Chang

In e-commerce, pricing strategies play a critical role in shaping consumer satisfaction. Since consumers rely on online review and rating to learn about products for making purchase decisions, retailers are under pressure to maintain a good record of rating. This study investigates the effect of pricing strategies and feedback mechanisms on consumer satisfaction in the context of online retail. We executed laboratory experiment with a 2x2x2 between subject factorial design, manipulating pricing strategy (partitioned vs. all-inclusive), feedback mechanism (reward vs. penalty framing), and delivery outcome (success vs. failure). The results of the study depict that when the carrier successfully delivered the product on time, consumers in the all-inclusive pricing strategy are always more satisfied with the service. When the carrier fails to deliver the product on time, free shipping is favorable to carrier in reward framing and partitioned pricing is favorable to carrier in penalty framing. The lack of transparency related to shipping fees in the all-inclusive pricing strategy may induce greater dissatisfaction to consumers, particularly when coupled with the penalty framing mechanism. Consumers feel a sense of unfairness and distrust, impacting their satisfaction negatively. This contrasts with those under partitioned pricing, who have a clearer understanding of the cost breakdown. These insights contribute to understanding the relationship between pricing strategies, feedback mechanisms, and consumer satisfaction in e-commerce.

Keywords: Pricing strategy, Framing feedback mechanism, Delivery outcome, Consumer satisfaction

5 - Fast and Frugal Heuristic Decisions for Pricing Truckload Delivery Services in Reverse Auctions

Michael Haughton, Wilfrid Laurier University, Waterloo, ON, Canada

Given the uncertainties and time pressures that carriers encounter in deciding the prices to charge for their truckload delivery services, fast and frugal heuristics (FFHs) represent a natural option for the pricing decision. Using behavior experiments and mathematical modeling, we examine the logic and efficacy of FFH-based pricing decisions. Our findings on matters such as how FFHs perform relative to mathematical optimization have practical implication for small carriers who are assessing whether their FFH-based decisions are good enough or whether to invest in pricing optimization tools.

6 - Identifying the Synergic Effect of Topics in Different Segments and Rate Prediction- Case Study of Physicians' Review

Reza Jazemi, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Ensieh Alidadiani

Online reviews wield significant influence over consumer decisions, shaping perceptions and choices across diverse sectors. But the extent to which online ratings convey information about service provider quality and impact consumers' decisions remains uncertain. This study aims to investigate the synergistic effect of topics within different segments and its influence on rating predictions, focusing on physicians' reviews sourced from various online platforms. The primary objectives include exploring topic variability across specialties, identifying synergic effects between topics, and examining their correlation with star ratings.

The research employs a robust methodology combining text analytics techniques with machine learning algorithms. Leveraging Latent Dirichlet Allocation (LDA) for topic extraction, the study categorizes reviews into common topics and segment-specific topics. Supervised learning methods are then utilized to prioritize topics within each specialty segment. Additionally, sentiment analysis and Generalized Random Forest are employed to detect interactions between topics and predict rating outcomes.

The study is uncovering insightful patterns regarding the dynamics of topics and their interactions within physician segments. Certain topics exhibit varying importance across specialties, with synergistic effects between topics influencing overall ratings. By evaluation of these relationships, the research offers practitioners action plans for service improvement and empowers prospective patients with informative review summaries in specific segments. Ultimately, the findings contribute to the optimization of decision-making processes within the healthcare industry.

7 - Fare Ladder Optimization: Airline Revenue Management in a Volatile Environment

Yan Chen, National University of Singapore, Singapore, Singapore, Changchun Liu, Gang Guo, Hong-ming Tan, Chung-piaw Teo

Airline fare ladder pricing is a strategy used by airlines to optimize revenue by offering various seat prices on the same flight, catering to different customer segments. Traditionally, this has relied on historical sales data and competitor pricing. However, the COVID-19 pandemic has disrupted the industry, prompting a need for new revenue management approaches.

This study presents a novel framework for ladder pricing in unstable environments, allowing airlines to adapt forecasting models to handle fluctuations better. The framework identifies changepoints in booking volumes and incorporates them into sales prediction and price elasticity assessments. Additionally, a pricing engine is developed to assist analysts in crafting optimized fare ladders.Collaborating with a leading airline, we conduct a large-scale field experiment, using the Difference-in-Differences strategy, assesses the impact on revenue and booking metrics. Overall, this approach addresses the fare optimization challenge amidst instability.

MD13

Summit - 333

Optimization in Supply Chain Network Design

Contributed Session

Chair: Ilya Jackson, MIT, Cambridge, MA, United States

1 - Revolutionize Cold Chain: An AI/ML Driven Approach to Overcome Capacity Shortages

Ilya Jackson, MIT, Cambridge, MA, United States, Jafar Namdar, Maria Jesus Saenz

This research investigates how Artificial Intelligence (AI) and Machine Learning (ML) forecasting methodologies can be leveraged for cold chain capacity planning, specifically utilizing Prophet and Seasonal Autoregressive Integrated Moving Average parametrized through grid search. In collaboration with Americold, the world’s second-largest refrigerated logistic service provider, the study explores the challenges and opportunities in applying AI/ML techniques to complex operations covering 385 customers and a capacity of 73,296 pallet positions. We train and test several AI/ML and traditional statistical models using extensive data for every customer over 3.5 years. Based on the results, MAPE of 5% was achieved on the whole site level, and SARIMA outperformed ML models in most cases. Next, we show that developing and applying a Customer Segmentation Matrix has enabled more accurate forecasting and planning across various customer segments, addressing the issue of forecasting inaccuracies. This approach effectively improves forecasting inaccuracies, underscoring the significance of tailoring AI/ML models for demand forecasting within the cold-chain industry. Ultimately, this research presents an AI-driven approach that transcends mere forecasting, offering a practical pathway to manage capacity in light of the constraints.

2 - Optimizing Closed-Loop Location-Inventory Problem Under Dynamic Demand

Sungjune Lee, Korea University Business School, Seoul, Korea, Republic of, Daeki Kim

This study examines the closed-loop supply chain, where distribution centers deliver, collect and reprocess returned products. Returnable transport items such as pallets are the representative, with the central operator playing a crucial role. Effective control over the location of distribution centers and inventory management is essential to fulfill this role. Initially formulated as a nonlinear mixed-integer problem, the model is reformulated into a conic quadratic mixed-integer program with valid cuts. We employ a scenario tree generated by Latin Hypercube Sampling and k-means clustering to address multi-period demand uncertainty and multi-stage stochastic programming is utilized to address different timing of strategic and operational decisions. Our analysis reveals that higher inventory costs complicate the problem. Inventory mediates the demand under the inventory ceiling constraint of distribution centers. The branch-and-cut process is beneficial under certain model characteristics, such as a smaller number of distribution candidates relative to customers. While the stability of the scenario tree generation is proved, the solution from the scenario tree and the deterministic model are compared and it is found that the dynamic demand does not affect the number of opened distribution centers as the demand could be handled by the inventory order amount.

3 - Matheuristic algorithms for the Inventory Routing Problem with Split Deliveries

Luca Bertazzi, University of Brescia, Brescia, Italy, Minh Nho Dinh, Demetrio Laganà

We propose new matheuristic algorithms for the Inventory Routing Problem with Split Deliveries, for both the Order-Up-to Level and the Maximum Level replenishment policies. The first matheuristic is based on the Capacitated Concentrator Location problem, while the second is based on a MIP model, in which all routes are built on the basis of a giant TSP tour serving all customers from the depot. These matheuristics are very fast with respect to the original model, and have a very good performance on average. The routes found by these matheuristics, together with other routes generated on the basis of different criteria, are then used as input data in a route-based formulation, able to improve the quality of the final solution on well-known benchmark instances.

4 - Machine learning algorithms for the design and operation of facilities in future logistics systems: A Survey

Dilhani Marasinghe, The University of Tennessee at Chattanooga, Chattanooga, TN, United States, Pubudu Jayasekara

This survey paper provides an exploration of the utilization of machine learning (ML) algorithms in the design and operation of facilities within logistics systems. By synthesizing a diverse range of research literature, academic papers, and industry reports, the survey offers a comprehensive overview of the various ML techniques utilized in optimizing facility design and operations for logistics purposes. Key areas covered include the application of ML algorithms for facility location optimization, layout design, capacity planning, and demand forecasting. The survey highlights successful implementations of ML algorithms in improving efficiency, reducing costs, enhancing decision-making, and increasing overall performance in logistics operations Ultimately, this study aims to provide insights into how ML can continue to revolutionize the design and operation of facilities in future logistics systems.

5 - Leveraging Graph Neural Networks for Enhanced Supply Chain Network Design and Optimization

Seyed Ahmad Torabzadeh, University at Buffalo, Buffalo, NY, United States, Sabrina Casucci, Robert Dell

This talk presents an innovative approach to dynamic supply chain network design by integrating Graph Neural Networks (GNNs) with optimization techniques. The proposed framework leverages GNNs to make decisions regarding facility location and allocation within the supply chain network. Using historical data, demand forecasts, and candidate locations for suppliers, distribution centers, and demand zone locations, the GNN generates recommendations for facility locations and assignments based on learned spatial relationships, demand patterns, and transportation costs. Based on the recommendations made by the GNN, the optimization model prescribes optimal strategies to fulfill different customers' demands and inventory management strategies in different layers of the supply chain network. We present results for several scenarios.

MD14

Summit - 334

ML/data-driven Decision-making in Finance

Invited Session

Finance

Chair: Ruixun Zhang, Peking University, Beijing, N/A

1 - A characterisation of cross-impact kernels

Mathieu Rosenbaum, Ecole Polytechnique, Mumbai, India

Trading a financial asset pushes its price as well as the prices of other assets, a phenomenon known as cross-impact. We consider a general class of kernel-based cross-impact models and investigate suitable parametrisations for trading purposes. We focus on kernels that guarantee that prices are martingales and anticipate future order flow (martingale admissible kernels) and those that ensure there is no possible price manipulation (no-statistical-arbitrage-admissible kernels). We determine the overlap between these two classes and provide formulas for calibration of cross-impact kernels on data. We illustrate our results using SP500 futures data. This is joint work with Mehdi Tomas.

2 - Inference Preferences from Behaviors

Haoyang Cao, Johns Hopkins University, Baltimore, MD, United States, Renyuan Xu, Zhengqi Wu, Samuel Cohen, Lukasz Szpruch

In 1964, Kalman raised the question of whether it is possible to recover quadratic cost function from an optimal linear policy. Since then, generations of economists have been studying viable ways to reveal meaningful utility functions from rational decisions. About three decades ago, computer scientists started to look at this field of inverse reinforcement learning (IRL), where algorithms are designed to effiiciently derive possible reward functions that have motivated an observed optimal policy; in the 2010's, Ziebart's "MaxEnt IRL" pushed this field to the spotlight, followed by GAIL, guided cost learning and many other influential IRL algorithms. However, there is always this nagging question of how to guarantee the IRL-identified reward function truely reflects the preferences of the demonstrating agent. This question becomes even more relevant today given the growing popularity of robo-advising: in order to provide suitable investment plans, the robo-advisor needs to learn about the clients' preferences of the investment strategies, the growth of wealth, and the timing. Such needs motivate us to re-examine IRL, its identifiability and the inference of preferences for an investor.

3 - Legislator Tweets about the Environment and the Returns of Green versus Brown Stocks

Milind Goel, London Business School, London, United Kingdom, Victor DeMiguel, Javier Gil-Bazo, Roberto Gomez-Cram

We study the impact of legislator tweets (X posts) that discuss environmental topics on the returns of green versus brown stocks. We use natural language processing (NLP) methods to identify tweets that discuss climate-related issues and to establish their sentiment towards green transition. We then study price changes of stocks in the minutes around each tweet to show that the prices of green (brown) stocks increase (decrease) in a statistically significant manner when legislators support green transition. Our high-frequency identification approach enables us to establish a causal relation between the opinions expressed in the tweets and the returns of green versus brown stocks. Our focus on legislator tweets suggests that the effect takes place via a regulatory channel. Progressively expanding the event horizon (over which we compute returns) from five minutes after the tweet to one day increases the economic relevance of our findings by an order of magnitude and reveals that it takes at least one day for the market to absorb the information contained in the tweets. Finally, we analyse the economic mechanism underlying the observed effects by studying how our findings depend on investor attention, institutional ownership, macroeconomic variables, the topic of the environmental tweet, and the legislator direct involvement in environmental regulation.

MD15

Summit - 335

Online Decision Making with Various Forms of Predictions

Invited Session

Revenue Management and Pricing

Chair: Vahideh Manshadi, Yale, Westwood, United States

Co-Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

1 - Magnolia: Matching Algorithms via Gnns for Online Value-to-Go Approximation

Anders Wikum, Stanford University, Stanford, CA, United States, Alexandre Hayderi, Amin Saberi, Ellen Vitercik

Online Bayesian bipartite matching is a central problem in digital marketplaces and exchanges, including advertising, crowdsourcing, ridesharing, and kidney exchange. We introduce a graph neural network (GNN) approach that emulates the combinatorially-complex optimal online algorithm. This optimal algorithm selects actions (e.g., which nodes to match) by computing each action's *value-to-go (VTG)*, which is the expected weight of the matching if the algorithm takes that action, and then chooses each subsequent action optimally. We train a GNN to estimate VTG and show empirically that this GNN returns high-weight matchings across a variety of tasks. Moreover, we identify a common family of graph distributions in spatial crowdsourcing applications, such as rideshare, under which VTG can be efficiently approximated by aggregating information within local neighborhoods in the graphs. This structure matches the local behavior of GNNs, providing theoretical justification for our approach.

2 - An Application of Online RL in Matching at Lyft

Mehdi Golari, Lyft, Seattle , WA, United States, Vijay Narasiman

To better match drivers to riders in our ridesharing application, we revised Lyft's core matching algorithm. We use a novel online reinforcement learning approach that estimates the future earnings of drivers in real time, and we use this information to find more efficient matches. This change was the first documented implementation of a ridesharing matching algorithm that can learn and improve in real time. We evaluated the new approach during weeks of switchback experimentation in most Lyft markets and estimated how it benefited drivers, riders, and the platform. In particular, it enabled our drivers to serve millions of additional riders each year, leading to more than $30 million per year in incremental revenue. Lyft rolled out the algorithm globally in 2021.

3 - Online Algorithms with Predictions

Debmalya Panigrahi, Duke University, Durham, NC, United States

In this talk, I will give an overview of the emerging area of online algorithms with predictions, where machine-learned predictions about the future are used to overcome information-theoretic barriers in online decision-making. I will describe the motivation behind the emergence of this research area, tell some of the early success stories, and outline some possible directions for further exploration in the near future.

4 - Best of Many in Both Worlds: Online Resource Allocation with Predictions Under Unknown Arrival Model 翻译 搜索 复制

Lin An, Carnegie Mellon University, Pittsburgh, PA, United States, Andrew Li, Benjamin Moseley, Gabriel Visotsky

Online decision-makers today can often obtain predictions on future variables, such as arrivals, demands, inventories, etc. These predictions can be generated from simple forecasting algorithms for uni-variate time-series, all the way to state-of-the-art machine learning models that leverage multiple time-series and additional feature information. However, the prediction quality is often unknown to decisions-makers a priori, hence blindly following the predictions can be harmful. In this talk, we address this problem by giving algorithms that take predictions as inputs and perform robustly against the unknown prediction quality.

We consider the online resource allocation problem, one of the most generic models in revenue management and online decision-making. In this problem, a decision-maker has a limited amount of resources, and requests arrive sequentially. For each request, the decision-maker needs to decide on an action, which generates some rewards and consumes some resources, without knowing the future requests. The decision-maker's objective is to maximize the total rewards subject to resource constraints. We take the shadow price of each resource as prediction. Prediction quality is naturally defined to be the l_1 distance between the prediction and the actual shadow price. Our main contribution is an algorithm which takes the prediction of unknown quality as an input, and achieves asymptotically optimal performance under both requests arrival models (stochastic and adversarial) without knowing the prediction quality and the arrival model beforehand. Our algorithm's performance matches the best achievable performance of any algorithm had the arrival models and the accuracy of the predictions been known.

MD16

Summit - 336

RMP for Online Retail

Invited Session

Revenue Management and Pricing

Chair: A. Serdar Şimşek, The University of Texas at Dallas, Richardson, TX, United States

1 - Algorithmic Self-Preferencing on E-Commerce Platforms: Evidence from JD.Com

Zihan Zhao, Washington University in St. Louis, St. Louis, MO, United States, Dennis Zhang, Renyu Zhang

E-commerce platforms that are both designers and participants of a marketplace, such as Amazon and JD.COM, might leverage recommendation algorithms to preferentially promote their own products, a phenomenon termed "algorithmic self-preferencing." In response to the increasing scrutiny from both regulatory bodies and the academic community, our study introduces a comprehensive framework to define, predict, and detect instances of algorithmic self-preferencing within such e-commerce platforms.

We define algorithmic self-preferencing as the platform's excessive promotion of its own products over equivalent ones sold by third-party sellers, and establish the necessary and sufficient conditions for such a self-preferencing phenomenon. Our theoretical model predicts that a self-preferencing algorithm is characterized by higher consumer traffic (i.e., the number of clicks) but a lower conversion rate for the platform-owned product.

We then leverage a large public dataset from JD.COM to empirically identify algorithmic self-preferencing. Utilizing coarsened exact matching for causal identification, our empirical findings reveal that platform-owned products receive 55.65% more clicks yet exhibit 23.77% lower conversion rates compared to their identical third-party counterparts. This disparity not only evidences algorithmic self-preferencing on JD.COM but also validates our theoretical predictions.

Our extensive robustness checks confirm that the presence and magnitude of algorithmic self-preferencing are robust with respect to different identification strategies and model specifications. Finally, we further demonstrate the generalizability of our framework with data from Amazon, revealing similar patterns of self-preferencing.

2 - Price-You-See vs. Price-You-Pay: Field Experiments for Event Ticket Pricing

Serkan Mehder, University of Colorado Boulder, Boulder, CO, United States, Maxime Cohen, Ovunc Yilmaz

Drip pricing, which involves advertising only a portion of a product's price with additional charges revealed later in the purchasing process, is becoming increasingly popular in various industries. Despite academic research showcasing the profitability of drip pricing, significant public debate surrounds the issue of price transparency. Partnering with an event ticket reseller, we conduct multiple field experiments and examine consumer behavior under various levels of concealed fees.

3 - Assortment and Fulfillment Optimization in Online Grocery Retail

Renjun Hu, University of Michigan, Ann Arbor, MI, United States, Hyun-Soo Ahn, Lennart Baardman

Despite the rapid growth of online grocery retail, most supermarkets have not turned this growth into profit. A key reason is the excessively high picking and delivery cost in an industry with low profit margins. Technological innovation, in the form of the Micro Fulfillment Center (MFC), has great potential to reduce costs by improving the efficiency in picking operations. The MFC is a small local warehouse, often attached to a brick-and-mortar store or regular fulfillment center, which efficiently fulfills a subset of the assortment. For a grocery retailer with an online store and an MFC, it is important to decide which items are offered in the assortment of the online store and which items are fulfilled from the MFC. This is especially important in online grocery retail where cart abandonment is ubiquitous. If customers cannot find the items that they are looking for, they can abandon their shopping cart and move to another online grocery retailer. In this work, we formulate the assortment and fulfillment problem for an online grocery retailer with an MFC whose customers exhibit cart abandonment behavior. We show that the problem is difficult to solve and that its parameters can be complicated to estimate. Hence, we reformulate the optimization problem such that its parameters are estimable with readily available data in the industry, and we develop an algorithm to quickly compute near-optimal decisions at large scale. Both in theory and computations, we can show that our algorithm generates good approximations.

4 - Blessing or Curse? An Investigation of Free-Return Services in E-Commerce

Ahmet Serdar Simsek, University of Texas at Dallas, Richardson, TX, United States, Jiarui (Gary) Wei, Ozalp Ozer

Third-party sellers on e-commerce marketplaces can choose to fulfill their orders through the marketplace’s fulfillment service , referred to as Fulfillment by Platform (FBP), or by managing their own fulfillment operations, referred to as Fulfillment by Seller (FBS). We investigate the effects of the free-return services offered by the FBP service on consumers’ purchase decisions and sellers’ revenues. To do so, we develop a stochastic model that captures the dynamics of consumer purchasing behavior and return decisions under the influence of the return policy. The leniency of return policies impacts product sales, return rates, and sellers' profits. On the one hand, a lenient return policy encourages consumers to purchase products with a reduced sense of risk, boosting overall sales. On the other hand, it increases return rates, potentially eroding sellers’ profitability. We examine how the relative profitability of FBP compared with FBS varies for retailers as product uncertainty and heterogeneity in consumers' valuation across the population change. We find that FBP is more profitable for sellers with products exhibiting high consumer heterogeneity or low product uncertainty. Our findings provide guidance to sellers on which fulfillment option is better suited for different types of products.

MD17

Summit - 337

New Trends in Discrete Choice Models

Invited Session

Revenue Management and Pricing

Chair: Gustavo Vulcano, Universidad Torcuato Di Tella, Ciudad Autónoma Buenos Aires, Argentina

1 - Revenue Management with Flexible Products

Huseyin Topaloglu, Cornell University, New York, NY, United States, Wenchang Zhu

We consider revenue management problems with flexible products, where we decide which set of resources to use to serve a customer during the service period at the end of the selling horizon. Flexible products allow firms to pool resource capacities and increase their resource utilization. We build on the earlier work on revenue management problems without flexible products to give a constant-factor approximation.

2 - Personalized Substitution Suggestions in Online Grocery Retailing

Luigi Laporte, UTDT - Universidad Torcuato Di Tella, Buenos Aires, Argentina, Srikanth Jagabathula, Daniel Corsten

Online grocery retailers have recently adopted presenting customers with personalized substitution suggestions during the ordering stage to avoid the costs associated with out-of-stocks. Selecting three products to offer as substitution suggestions to maximize the probability of the customer selecting one is a difficult problem because assortments are large, customer data are very sparse, and product availability must be considered both to infer customer preferences from past transactions and to decide which suggestions to present, accounting for the lag between the ordering stage and the picking stage in the online retail setting. We investigate the application of discrete choice models to the problem. We develop and test a three-component framework in which a ranking-generating component is followed by a relevancy filter and a operational filter. We tested the prediction power of our ranking-generating component -- based on a partial order, DAG-based choice model -- and found that it outperforms MNL and LC-MNL choice models in an extensive numerical experiment with offline data. We performed a large-scale field experiment with a European online grocery retailer. The results show that the relevancy filter is essential to counterbalance the unobservability of the customers' consideration sets. They also show that our proposed framework outperforms a machine learning-based model and that the performance improvement is more pronounced for customers with few transactions in the focal category. Our study results indicate that retailers can employ our proposed framework in practice as a general approach to provide personalized substitution suggestions.

3 - Exact Logit-Based Product Design

Velibor Misic, UCLA Anderson School of Management, Los Angeles, CA, United States, Irem Akcakus

The share-of-choice product design (SOCPD) problem is to find the product, as defined by its attributes, that maximizes market share arising from a collection of customer types or segments. When customers follow a logit model of choice, the market share is given by a weighted sum of logistic probabilities, leading to the logit-based share-of-choice product design problem. In this paper, we develop a methodology for solving this problem to provable optimality. We first analyze the complexity of this problem, and show that this problem is theoretically intractable: it is NP-Hard to solve exactly, even when there are only two customer types, and it is furthermore NP-Hard to approximate to within a non-trivial factor. Motivated by the difficulty of this problem, we propose three different mixed-integer exponential cone programs of increasing strength for solving the problem exactly, which allow us to leverage modern integer conic program solvers such as Mosek. Using both synthetic problem instances and instances derived from real conjoint data sets, we show that our methodology can solve large instances to provable optimality or near optimality in operationally feasible time frames and yields solutions that generally achieve higher market share than previously proposed heuristics.

4 - The Mallows Model of Discrete Choice in Operational Contexts

Gustavo Vulcano, Universidad Torcuato Di Tella, Ciudad Autónoma Buenos Aires, Argentina, Srikanth Jagabathula

We study the Mallows discrete choice model, which is well-known in the machine learning community, and novel in the operations management community. .

We first introduce the Mallows model and present some well known results in the literature inferred from the rank-based definition of the model. Then, we derive some new results related to the computation of Mallows probabilities over partial order structures, which generalize the results over rank lists. These new results are fundamental to develop different estimation algorithms to calibrate a Mallows model from anonymous data. We present three variants of the estimation algorithm founded on different ways to envision the input data. Finally, we show how to extend the estimation method to the case of preferences described by partial orders.

Our numerical experiments on synthetic transaction data and real-world panel data identify frameworks where the model allows more accurate predictions compared to state-of-the-art alternative methods.

MD18

Summit - 338

Advances in Experimentation and Policy Learning in Managerial Settings

Invited Session

Revenue Management and Pricing

Chair: Spyros Zoumpoulis, INSEAD, Fontainebleau, France

Co-Chair: Stefanos Poulidis, INSEAD, Fontainebleau, 77300, France

1 - Policy Evaluation and Policy Learning Under the F-Sensitivity Model

Zhimei Ren, University of Pennsylvania, Philadelphia, PA, United States, Ying Jin, Sharon Wang, Ruohan Zhan, Zhengyuan Zhou

In this talk, I will introduce the f-sensitivity model, a new sensitivity model that characterizes the violation of unconfoundedness in causal inference. It assumes the selection bias due to unmeasured confounding is bounded “on average”; compared with the widely used point-wise sensitivity models in the literature, it is able to capture the strength of unmeasured confounding by not only its magnitude but also the chance of encountering such a magnitude.

Then, I will talk about policy evaluation and policy learning under our new model based on a distributional robustness perspective. For policy evaluation, our estimators are asymptotically normal under mild conditions. We also propose a learning algorithm returning the policy that maximizes the policy value estimator, for which we provide theoretical guarantees on the sub-optimality gap. The proposed methods are implemented and evaluated in numerical simulations, demonstrating substantial improvement compared with existing benchmarks.

2 - Detecting Service Slowdown Using Observational Data

Kuang Xu, Stanford Graduate School of Business, Stanford, CA, United States, Gal Mendelson

Being able to detect service slowdowns is crucial to many operational problems. We study how to use observational congestion data to detect service slowdown in a multi-server system, and in particular, the statistical implications of running adaptive congestion control mechanisms in such settings. We show that a commonly used summary statistic that relies on the marginal congestion measured at individual servers can be highly inaccurate the presence of adaptive congestion control. We propose a new statistic based on potential routing actions, and show it provides a much more robust signal for server slowdown in these settings. Unlike the marginal statistic, potential action aims to detect changes in the routing actions, and is able to uncover slowdowns even when they do not reflect in marginal congestion. Our work highlights the complexity in performing observational statistical analysis for service systems in the presence of adaptive congestion control. Our results also suggest that practitioners may want to combine multiple, orthogonal statistics to achieve reliable slowdown detection.

3 - Estimating Treatment Effects Under Recommender Interference: a Structured Neural Networks Approach

Ruohan Zhan, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Shichao Han, Yuchen Hu, Zhenling Jiang

Recommender systems are essential for content-sharing platforms by curating personalized content. To evaluate updates of recommender systems targeting content creators, platforms frequently engage in creator-side randomized experiments to estimate treatment effect, defined as the difference in outcomes when a new (vs. the status quo) algorithm is deployed on the platform. We show that the standard difference-in-mean estimator can lead to a biased treatment effect estimate. This bias arises because of recommender interference, which occurs when treated and control creators compete for exposure through the recommender system. We propose a “recommender choice model” that captures how an item is chosen among a pool comprised of both treated and control content items. By combining a structural choice model with neural networks, the framework directly models the interference pathway in a microfounded way while accounting for rich viewer-content heterogeneity. Using the model, we construct an estimator of the treatment effect that is consistent and asymptotically normal. We demonstrate its empirical performance with a field experiment on Weixin short-video platform. Besides the standard creator-side experiment, we carry out a costly blocked double-sided randomization design to obtain a benchmark estimate without interference bias. We show that the proposed estimator significantly reduces the bias in treatment effect estimates compared to the standard difference-in-mean estimator.

4 - Vague Price Optimization

Wen yun, columbia university , new york , NY, United States, Assaf Zeevi

Price experimentation is increasingly used by firms across different industries and has become an integral part of revenue management practices. Because of sensitivity and potential impact on customer relationship management, prices are often set and modified by pricing specialists in a manual manner driven by various heuristics. We report on work with a large global manufacturing company that focuses on the design of a data-driven "price recommendation" engine. This engine provides pricing specialists with "vague" pricing suggestions: the objective is not to precisely identify the "best" action but rather to provide a recommendation that also incorporates and is driven by real world practices and constraints. We demonstrate the system's effectiveness when pricing specialists adhere to the provided suggestions, and investigate the losses associated with various forms of noncompliance. This analysis offers insights into possible trade-offs between practical constraints and algorithmic design principles and will be illustrated on synthetic as well as real sales data.

MD19

Summit - 339

Strategic Dynamics and Optimization in Crowdfunding and Start-Ups

Contributed Session

Chair: Dowon Lee, University at Buffalo, The State University of New York, Williamsville, NY, United States

1 - Effect of a Wholesale Price Contract on the risk of Delivery Failure in a Crowdfunding Campaign.

Joyaditya Laik, Bucknell University, Lewisburg, PA, United States, Qian Zhang, Nabita Penmetsa

We study how a traditional wholesale price contract affects the optimal target and pledge amount in a crowdfunding campaign, when an entrepreneur depends on his supplier for components. A crowdfunding campaign is a verifiable signal that a supplier may use to gauge the demand of the product and adjust her wholesale price, in response to the prices the entrepreneur sets for a product both in the crowdfunding campaign and post-campaign market. We find that the efficiency of the crowdfunding market, in predicting post-crowdfunding demand affects the parameters set in this transaction. We contribute to the less studied field of a nascent supply chain, where the entrepreneur is in a weaker bargaining position, and therefore depends on the outcome of market signals, like crowdfunding, to provide evidence about the future prospect of the product. Through this paper we model the benefits that an entrepreneur gets from engaging in a crowdfunding campaign, by using the subscription levels to set prices. Furthermore, our study also shows that the target and pledge amounts are a consequence of contracts set with the suppliers.

2 - The Time-Varying Effects of Knowledge Diffusion in Open Source Projects

Orcun Temizkan, Ozyegin University, Istanbul, Turkey, Ram Kumar

Knowledge diffusion among software developers facilitates knowledge creation in Open Source (OS) projects. Prior research has illustrated that connectivity between software developers is related to knowledge creation in OS projects. However, the effect of knowledge diffusion over time is under researched. We illustrate that knowledge diffusion among software developers depends on the connectivity of software developers within projects and that the effect of developer connectivity on knowledge creation is time-dependent. Data collected from real-world OS projects is used in order to illustrate these effects. Research and managerial implications will also be discussed.

3 - Managing Start-up Growth Before Facility Investments

Yuanyuan Li, ESSEC Business School, Cergy, France, Riccardo Mogre, Felix Papier

We study the growth of the startup before it purchases its own plant. With limited capital, startups often leverage contract manufactures until they gain enough traction to secure facility investments. Herein lies a key question: whether to grow rapidly to buy the plant earlier, risking excessive capacity adjustment costs with contract manufacturers, or to grow slowly and then acquire the plant later. To find the tradeoff, we first analyze a stochastic model with the stochasticity in the number of new growths generated by sales efforts. Then we analyze a static convex model, assuming a deterministic number of growths resulting from a certain sales effort, where we derive closed-form solutions for the optimal sales effort in each period and the optimal number of periods. This study provides insights for startups navigating the path to facility investment while optimizing their growth strategies.

4 - Optimal Product Launch and Innovation Investment Decisions for a Niche Firm

Akhil Singla, Northwestern University, Evanston, IL, United States, Izak Duenyas, Seyed Iravani, Jacqueline Lane

We consider the product launch and innovation investment decisions for a niche firm that produces a single product in the presence of a multi-product rival firm. The rival continuously launches its new product generations, which increases its competitive advantage compared to the niche product, consequently reducing the niche firm's revenue. However, to gain higher revenues, the niche firm may launch its R&D product into the market by incurring a launch cost or, to gain competitive advantage against the rival, the niche firm may innovate by investing in R&D. The novel aspect of this problem is that the niche firm’s product is an auxiliary or peripheral part of a multi-product firm’s business; therefore, only the niche firm views the multi-product firm as a direct competitor. We develop a MDP model to analyze the niche firm’s decisions on when to launch and how to innovate, incorporating stochastic innovation gains. We characterize the structure of the optimal launch decision policy and study its sensitivity with respect to model parameters. Due to non-monotonicity in the structure of the optimal policy driven by the innovation investment decisions, we partially characterize the optimal policy that informs product launch and innovation investment decisions. Therefore, we propose a triple-threshold heuristic policy that performs very close to the optimal solution and provides a simpler decision rule for the product launch and innovation investment decisions. We identify the conditions under which the commonly used policy to launch the product after fixed time intervals can incur huge costs for the firm.

5 - Scheduling ground-to-satellite communications with cislunar orbits and interception risk

Dowon Lee, University at Buffalo, Amherst, NY, United States, Moises Sudit, Christopher Nebelecky, Jeremy Dening, Chase Murray

The Moon has gained significant attention recently as indicated by NASA's Artemis missions, with the goal of establishing a long-term presence for exploration. Satellites in lunar or cislunar orbit will be necessary to enable communications from the surface of the Moon. The scheduling of communication requests on ground stations to satellites will need to be optimized. The scheduling of ground to satellite communications to maximize the profit from scheduled requests is the Satellite Range Scheduling Problem. However, it is also of interest to ensure the communications are secure. When another satellite is also within the cone of visibility during a transmission, there is a risk that the transmission is intercepted. With the goal of secure communications, the risk from scheduled requests may need to be limited or minimized. Two discrete time integer programming models are introduced to solve the satellite range scheduling problem with different perspectives: (1) maximize profit with a risk limit, and (2) minimize risk with a profit requirement. An analysis on the performance is conducted using a variety of lunar and cislunar orbit combinations. This work is funded by the AFOSR through the Space University Research Initiative (SURI).

MD20

Summit - 340

Decision Analysis Through Information Sharing

Invited Session

Decision Analysis Society

Chair: Behnam Malmir, Georgia Southern University, Savannah, GA, United States

Co-Chair: Aram Bahrini, University of Illinois Urbana-Champaign, Champaign, IL, United States

1 - Online Opinion Leadership during Crises

Behnam Malmir, Georgia Southern University, Savannah, GA, United States

During pandemics like COVID-19, public health officials face challenges disseminating guidance. Opinion leaders (OLs) serve as trusted sources, aiding guideline adoption. This study presents the Pandemic OL Analyzer (POLA), identifying OLs on Twitter. Trust emerges as the most influential trait, followed by expertise, uniqueness, innovation, and reputation. OLs include organizational and individual types, with organizational ones exerting more influence. OL activity spans eight domains: government planning, blood issues, reports, public actions, organizational aid, vaccination, community solutions, and general health. POLA aids in efficient OL identification for effective pandemic communication.

2 - Information Sharing in Closed Loop Supply Chains

Aram Bahrini, University of Illinois Urbana-Champaign, Champaign, IL, United States, Robert J. Riggs, Maryam Esmaeili

In this study, we explore how information sharing affects the performance of a two-level closed-loop supply chain consisting of one manufacturer and one retailer, where the manufacturer is the leader and the retailer is the follower. The manufacturer sells products through the retailer in the forward supply chain and collects used products in the reverse supply chain. Despite demand uncertainty, both manufacturers and retailers possess private information about demand. We determine the Stackelberg equilibrium for retail price, wholesale price, and the rate of collecting used products. We then compare their profits under scenarios with and without information sharing. Our findings reveal that information sharing in this leader-follower closed-loop supply chain model does not always enhance the profits of both the manufacturer and the retailer.

3 - Irrigation Decisions in a Game-Theoretic Landscape: Strategic Risk Dominance and Topology Design

Mobin Zarreh, Arizona State University, Tempe , AZ, United States, Paul Grogan

This research investigates the pivotal role of constructed water infrastructures in enhancing societal well-being and economic development, with a specific focus on the design and user dynamics of infrastructure systems. The study centers on flood irrigation systems managed by the Salt River Project (SRP) in Arizona—a communal infrastructure that enables both utility customers and homeowners to efficiently irrigate crops and lawns through an extensive canal network. This investigation treats system design as a bi-level problem featuring three asymmetric decision-makers engaged in a strategic design game. It incorporates metrics of risk dominance and operational choices to deepen the understanding of strategic interactions and their effects on both the resilience of the infrastructure and the benefits to users. Through the modeling of various scenarios, this study examines differing infrastructure topologies and their impact on collaborative decision-making processes. It specifically contrasts collaborative flood irrigation designs with those using city (municipal) water resources for irrigation. An agent-based model (ABM) is developed to simulate these scenarios, providing insights into the interplay between infrastructure design and user strategies, and assessing the robustness of collaborative designs. This methodological framework not only identifies optimal design and operational strategies but also supports policy development aimed at ensuring the sustainability and long-term viability of water infrastructure systems, thus promoting sustainable development, and enhancing quality of life.

4 - Design of a Distribution Network in a Multi-product, Multi-period Green Supply Chain System Under Demand Uncertainty

Azam Boskabadi, Washington State University, Pullman, WA, United States, Mirpouya Mirmozaffari

This paper proposes a novel fuzzy mathematical model for a distribution network design problem in a multi-product, multi-period, multi-echelon, multi-plant, multi-retailer, multi-mode of transportation green supply chain system. The three purposes of the model are to minimise total network cost, maximise net profit per capita for each human resource, and diminish CO2 emission throughout the network. P-hub median location with multiple allocations is used for locating the distribution centres. One scenario is designed for fuzzy customer demands with a trapezoidal membership function. Furthermore, the model determines the design of the network (selecting the optimum numbers, locations of plants, and distribution centres to open), finding the best strategy for material transportation through the network with the availability of different transportation modes, the capacities level of the facilities (plants or distribution centres (DCs)), and the number of outsourced products. Finally, all uncertain customer demands for all product types can be satisfied based on the methods mentioned above. This multi-objective mixed-integer non-linear mathematical model is solved by NSGA-II, MOPSO and a hybrid meta-heuristic algorithm. The results show that NSGA-II is the exclusive algorithm that obtains the best result according to the evaluation criteria.

5 - Intervention in Collaborative System Design to Increase Efficiency: A Communication Channel for Technical and Social Information Exchange

Alkim Avsar, Arizona State University, Tempe, AZ, United States, Paul Grogan

Collaborative systems involve multiple actors pooling their resources and efforts to achieve a complex goal that surpasses the capabilities of any single entity. Distributed Satellite Systems (DSS) exemplify this, where numerous spacecraft collaborate with diverse functionalities to fulfill a shared objective. However, while collaborations offer promising outcomes, their failure can lead to substantial sunk costs for involved parties. In such systems, actors often prioritize their self-interest, selectively sharing or withholding information, resulting in reduced transparency and lack of control over decision variables, leading actors to make strategic decisions. Consequently, risk emerges from both technical and social sources, making actor interactions critical. Facilitating effective communication among actors and providing a more transparent technical picture might help actors make more effective strategic decisions. This study proposes an intervention: a *system mediator*, to enhance information exchange among actors and provide a more transparent technical picture within collaborative systems. The study conducts a human experiment with a bi-level collaborative system problem to assess the impact of the system mediator. In the experimental process, designers represent decision-making authorities of different product design companies and engage in a design and strategic decision-making process. Experimental groups, with and without the system mediator, are compared to evaluate its efficacy. Results indicate that the system mediator significantly enhances overall system efficiency by augmenting actors' access to technical and social information. Understanding the mechanisms underlying this intervention's success and refining the system mediator can facilitate the development of management strategies conducive to optimizing complex collaborative systems.

MD21

Summit - 341

Making Army Installations More Resilient

Invited Session

Decision Analysis Society

Chair: Gregory Parnell, University of Arkansas, Fayetteville, AR, 72702, United States

1 - Leveraging Data and Artificial Intelligence for Supporting Installation Planning and Resilience Strategies

Randy Buchanan, USACE - ERDC, Vicksburg, MS, United States

An all-hazards risk approach to military installation planning increasingly requires a significant effort toward understanding climate change impacts. Climate change threatens physical and operational security of U.S. military bases around the world. Threats include sea level rise, droughts, wildfires, floods, and numerous extreme weather events, with each installation possessing unique vulnerabilities. Weather and climate intelligence efforts support the modernization of installation decision-making processes by applying complex computational analytics and in some cases high-performance computing assets to inform decision makers for making real-time holistic weather-related decisions as well as supporting long-term climate resilience planning. This presentation discusses these hazards and their impacts, and the return on investment for resilient strategies, as well as emerging capabilities such as artificial intelligence being leveraged to support large historical and real-time data set analytics.

2 - Incorporating Military Installation Resilience Plan (MIRP) Resilience and Risks/hazards into Installation Planning

Mary Mitchell, US Army Information Technology Laboratory, Vicksburg, MS, United States, Su Wolters, Tim Perkins, Shelia Barnett, Matt Swanson, John Richards

Military installations need to modernize to mitigate threats from increased complexity and potential risks. Central to future military installations is revolutionizing installation master planning. Installation master planning is the process to plan out infrastructure projects and prepare for upcoming mission requirements and risks, mandated by UFC 2-100-01. PLANNER is an Army Installation Modernization Pilot Program (AIMP2) project managed by ERDC under leadership of the Assistant Secretary of the Army for Installations Energy & Environment's Strategic Integration (ASA-IE&E(SI)) and powered by the Virtual Toolbox for Installation Mission Effectiveness (VTIME). Through the data lake provided by VTIME, PLANNER digitizes and operationalizes integrated installation planning, allowing the Army to overcome information stovepipes, and move away from late-to-need, time-intensive static plans developed at high-cost. As weather and climate threats impact installations, and impacts are forecasted to continue to increase, a key component of transforming installation planning includes an all-hazards approach to risk analysis to form clearer pictures of individual installations' resilience. Our team conducted a Business Process Reengineering (BPR) study to understand the processes, people, policies, technology, and information necessary for establishing All-Hazards-Risk-Analysis (AHRA) and deliver it as a PLANNER component of VTIME. The AHRA component will assist with incorporating resilience, risks, and hazards and linking them to resilience strategies for integrated installation resilience planning process and enable continuous risk/deficiency assessments to allow installations to better prepare for potential threats and hazards. This presentation will focus on the development of the AHRA and its potential to connect risks and hazards to the installation planning process.

3 - ROI Model to Evaluate Options to Make Installations Resilient to Extreme Weather

Gregory Parnell, University of Arkansas, Fayetteville, AR, United States

As the Department of Defense recognizes the potential for severe weather events caused by climate change, they plan to develop options to mitigate the impact on installation facilities and reduce the disruption in operations. While the Army recognizes that climate change will require adaptation costs, the data surrounding imposed costs to installations has not been integrated across the department. We are working to support the HQDA and G-9 by developing a Return-on-Investment (ROI) model for PLANNER (the Army's new data-driven master planning tool), to inform cost-effective modernization of installation infrastructure to make the facilities more resilient to the impacts of extreme weather.

4 - LLM-assisted Support for Infrastructure and Installation Resilience Planning

George Gallarno, US Army Engineer Research and Development Center, Vicksburg, MS, United States, Jaylen Hopson

The quick retrieval of information is important for informing consistent, transparent, and data-driven decision making during the development and execution of an installation resilience plan. Specifically, being able to reference standards, guidelines, and regulations is necessary to develop and execute an appropriate infrastructure strategy. In this study a large language model (LLM) is enhanced with domain-specific knowledge to quickly provide information to the end user on U.S. Army Corps of Engineers construction management-related topics. Specifically, retrieval augment generation and prompt engineering techniques were used to develop and deploy a proof-of-concept LLM using cloud-based tools. This proof-of-concept capability provided at least a 20x time savings for the test user group, helping provide quick reference to many applicable standards, guidelines, and regulations related to their various queries. This research shows the applicability of LLMs to the field of resilience planning by improving the quality and consistency of documentation as well as, more generally, enhancing the productivity of planning-related personnel.

This presentation will overview the methodology used to develop and deploy the proof-of-concept LLM, discuss Department of Defense-related constraints and limitations, and summarize the lessons learned from deployment of the proof-of-concept capability. Applications to broader infrastructure resilience planning and decision support will also be discussed.

5 - AI for Installation Decision Support

Patrick Ables, US Army Information Technology Laboratory, Vicksburg, MS, United States, George Gallarno, Randy Buchanan

This presentation delves into the transformative potential of artificial intelligence (AI) in enhancing decision-making processes within government installations across various domains. As an administrative tool, AI demonstrates exceptional effectiveness in aiding decision-making by automating and optimizing tasks such as summarizing internal documentation, scheduling, and predictive maintenance. AI-powered natural language processing (NLP) techniques streamline decision-making by automating the summarization of extensive internal documentation, extracting key insights from reports, policy documents, and meeting minutes, facilitating efficient information retrieval and trend identification enabling decision-makers to prioritize critical issues effectively. AI also plays a crucial role in analyzing vast datasets to uncover patterns and predict future conditions, thereby informing strategic decisions related to other administrative functions. For example, in environmental management, AI algorithms can analyze weather and climate data to inform infrastructure planning and disaster preparedness. More broadly, AI optimizes maintenance schedules by leveraging predictive analytics to forecast equipment failures based on historical performance data, usage patterns, and environmental factors. By implementing proactive maintenance strategies, government installations minimize downtime, reduce maintenance costs, and prolong asset lifespan. Through the integration of AI technologies, government installations can enhance decision-making processes, improve operational efficiency, and mitigate risks across diverse operational domains. By harnessing the power of AI to analyze data, extract insights, and provide actionable recommendations, decision-makers can navigate complexities, allocate resources effectively, and achieve their organizational objectives with greater agility and foresight, ultimately benefiting citizens and stakeholders.

MD22

Summit - 342

Integrating Generative AI with Sequential Decision-Making: Theoretical Advances and Practical Applications (II)

Invited Session

Decision Analysis Society

Chair: Yingru Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Ming Yin, Princeton University, Princeton, NJ, United States

Co-Chair: Ming Yin

1 - Efficient Reductions for Reinforcement Learning from Human Feedback

Gokul Swamy, Carnegie Mellon University, Pittsburgh, PA, United States

RL from human feedback (RLHF), also known as preference-based RL (PbRL), has emerged as a critical component of the fine-tuning procedure for large language models (LLMs). The standard recipe for RLHF involves first fitting a reward model from preference data before optimizing the learned reward via some RL algorithm. However, such an approach makes a key assumption: that the preferences we observe exhibit none of the intransitivity that frequently occurs in practice when aggregating preferences across a diverse population of raters. In this talk, we will discuss how taking a game-theoretic perspective on this problem (https://arxiv.org/abs/2401.04056) allows us to design efficient reduction-based algorithms that handle complex, intransitive preferences in a robust manner, match (and sometimes supersede) the performance of traditional approaches under idealized conditions, and elide reward modeling as well as adversarial training.

2 - Adapting Foundation Models Using Reinforcement Learning and Imitation Learning

Joey Hejna, Stanford University, Stanford, CA, United States

Modern foundation models are often trained to simply mimic a distribution of data. Such approaches, however, cause models to exhibit the same sub-optimalities present in their training dataset, which is particularly detrimental in sequential decision making settings. In this talk I will discuss how ideas from reinforcement learning an imitation learning can be used to mitigate this problem. First, I will discuss how we can use human feedback to directly optimize sequential policies, instead of myopically considering a bandits setting as usually done in LLMs. Next, I will show how ideas in inverse reinforcement learning can be used to quickly adapt models in the low-data regime. Finally, I will discuss how similar insights can transfer to foundation models in the robotics domain, and discuss our recent progress in this area.

3 - Crispr-GPT: An LLM Agent for Automated Design of Gene-Editing Experiments

Kaixuan Huang, Princeton University, Princeton, NJ, United States, Yuanhao Qu, Henry Cousins, William Johnson, Di Yin, Mihir Shah, Denny Zhou, Russ Altman, Mengdi Wang, Le Cong

The introduction of genome engineering technology has transformed biomedical research, making it possible to make precise changes to genetic information. However, creating an efficient gene-editing system requires a deep understanding of CRISPR technology, and the complex experimental systems under investigation. While Large Language Models (LLMs) have shown promise in various tasks, they often lack specific knowledge and struggle to accurately solve biological design problems. In this work, we introduce CRISPR-GPT, an LLM agent augmented with domain knowledge and external tools to automate and enhance the design process of CRISPR-based gene-editing experiments. CRISPR-GPT leverages the reasoning ability of LLMs to facilitate the process of selecting CRISPR systems, designing guide RNAs, recommending cellular delivery methods, drafting protocols, and designing validation experiments to confirm editing outcomes. We showcase the potential of CRISPR-GPT for assisting non-expert researchers with gene-editing experiments from scratch and validate the agent's effectiveness in a real-world use case. Furthermore, we explore the ethical and regulatory considerations associated with automated gene-editing design, highlighting the need for responsible and transparent use of these tools. Our work aims to bridge the gap between beginner biological researchers and CRISPR genome engineering techniques, and demonstrate the potential of LLM agents in facilitating complex biological discovery tasks.

4 - Do LLM Agent have regret?

Chanwoo Park, MIT, Cambridge, MA, United States, Xiangyu Liu, Asuman Ozdaglar, Kaiqing Zhang

Large language models (LLMs) have been increasingly employed for (interactive) decision-making, via the development of LLM-based autonomous agents. Despite their emerging successes, the performance of LLM agents in decision-making has not been fully investigated through quantitative metrics, especially in the multi-agent setting when they interact with each other, a typical scenario in real-world LLM-agent applications. To better understand the limits of LLM agents in these interactive environments, we propose to study their interactions in benchmark decision-making settings in online learning and game theory, through the performance metric of \emph{regret}. We first empirically study the {no-regret} behaviors of LLMs in canonical (non-stationary) online learning problems, as well as the emergence of equilibria when LLM agents interact through playing repeated games. We then provide some theoretical insights into the no-regret behaviors of LLM agents, under certain assumptions on the supervised pre-training and the rationality model of human decision-makers who generate the data. Notably, we also identify (simple) cases where advanced LLMs such as GPT-4 fail to be no-regret. To promote the no-regret behaviors, we propose a novel \emph{unsupervised} training loss of \emph{regret-loss}, which, in contrast to the supervised pre-training loss, does not require the labels of (optimal) actions. We then establish the statistical guarantee of generalization bound for regret-loss minimization, followed by the optimization guarantee that minimizing such a loss may automatically lead to known no-regret learning algorithms. Our further experiments demonstrate the effectiveness of our regret-loss, especially in addressing the above ``regrettable'' cases.

5 - Reinforcement Learning from Human Feedback: From Theory to Algorithm

Wei Xiong, UIUC, Urbana, IL, United States

Reinforcement Learning from Human Feedback (RLHF) is the leading technique to align foundation large language (LLM) model with the human preferences, and achieves tremendous successes in the application of Chat-GPT, Gemini, and Claude. We consider a standard mathematical formulation, the reverse-KL regularized contextual bandit for RLHF. Despite its widespread practical application, a rigorous theoretical analysis of this formulation remains open. In this presentation, we first establish the mathematical foundation of RLHF by proposing statistically efficient algorithms, which enjoy finite-sample guarantee under standard assumptions. We then move toward practical applications and show that the proposed algorithms can be practically implemented and empirically outperform existing strong baselines like DPO and RSO in real-world LLM experiments.

MD23

Summit - 343

On the Fairness for Optimization and Learning

Contributed Session

Chair: Eklavya Sharma, ISE, UIUC, Urbana, IL, United States

1 - Improving Approximation Guarantees for Maximin Share

Eklavya Sharma, ISE, UIUC, Urbana, IL, United States, Hannaneh Akrami, Jugal Garg, Setareh Taki

We consider fair division of a set of indivisible goods among n agents with additive valuations using the fairness notion of maximin share (MMS). However, since MMS allocations do not always exist [Kurokawa et al., JACM'18], the focus shifted to investigating its multiplicative approximations. Here the goal is to show the existence of α-MMS allocations (for the largest possible $\alpha < 1$), i.e, allocations that guarantee each agent at least α times her MMS value. A series of works in the last decade led to the state-of-the-art factor of $\alpha = 3/4 + 3/3836$ [Akrami and Garg, SODA'24].

We introduce a general framework of approximate MMS with agent priority ranking. We order the agents, and agents earlier in the order are considered more important. An allocation is said to be T-MMS, for a non-increasing sequence $T := (\tau_1, \ldots, \tau_n)$ of numbers, if the agent at rank i in the order gets a bundle of value at least τ_i times her MMS value. We show the existence of T-MMS allocations where $\tau_i \geq \max(3/4 + 1/12n, 2n/(2n+i-1))$ for all i. Furthermore, by ordering the agents randomly, we can get allocations that are (3/4 + 1/12n)-MMS ex-post and (0.8253 + 1/36n)-MMS ex-ante. We also investigate the limitations of our algorithm and show that it does not give better than (0.8631 + 1/2n)-MMS ex-ante

2 - Trading-off price for data quality to achieve fair online allocation

Vianney Perchet, ENSAE, Palaiseau, France, Patrick Loiseau, Nicolas Gast, Mathieu Molina

We consider the problem of online allocation subject to a long-term fairness penalty. Contrary to existing works, however, we do not assume that the decision-maker observes the protected attributes -- which is often unrealistic in practice. Instead they can purchase data that help estimate them from sources of different quality; and hence reduce the fairness penalty at some cost. We model this problem as a multi-armed bandit problem where each arm corresponds to the choice of a data source, coupled with the online allocation problem. We propose an algorithm that jointly solves both problems and show that it has a regret bounded by O(T^{1/2}). A key difficulty is that the rewards received by selecting a source are correlated by the fairness penalty, which leads to a need for randomization (despite a stochastic setting). Our algorithm takes into account contextual information available before the source selection, and can adapt to many different fairness notions. We also show that in some instances, the estimates used can be learned on the fly.

3 - Symmetrically Fair Allocations of Indivisible Goods

Connor Johnston, University of Florida, Gainesville, FL, United States, Aleksandr Kazachkov

We consider the allocation of indivisible goods with provable fairness guarantees which are satisfied regardless of which bundle of items each agent receives. One such fairness guarantee is envy freeness up to one good (EF1), under which an agent is permitted to envy another agent's bundle, but the amount of envy is no more than the value of the agent's favorite good in the envied bundle. When this fairness criteria is satisfied regardless of which bundle an agent receives, we call this allocation symmetrically envy free up to one good or symEF1. We discuss theoretical results about the existence of symEF1 allocations and computational results related to the frequency at which symEF1 allocations exist and model the problem as an integer program to verify if a symEF1 allocation exists. The main theoretical result of our paper is for 2 agents, there always exists a symEF1 allocation. This theorem is proved with a sufficient condition for the existence of symEF1 allocations which relates the valuation matrix to a graph. If this graph is n-colorable, then there exists a symEF1 allocation. We solve the integer program if our greedy algorithm cannot find a feasible allocation. The percentage of cases that needed the IP ranged from 7% to 99%. Our computational testing for cases of 3, 4, and 5 agents involved varying the number of items. Our results show that the proportion of allocations that satisfy the symEF1 constraint quickly tends to 1 with more items but it is not monotonically increasing.

4 - On-Demand Delivery: Estimation, Routing and Last Mile Solutions

Tayo Fabusuyi, University of Michigan, Ann Arbor, MI, United States, Majid Mirzanezhad, Yangyang Wang

This research develops a nationally scalable approach that estimates and forecasts the volume of on-demand delivery at fine geographic scales and implements an optimization model that fulfills the delivery with a focus on last mile solutions. Utilizing public use microdata sample files, we estimate demand using extreme gradient boosting and generate representative demand profiles through population synthesis. Additionally, we employ nearest-neighbor matching to fill data gaps and maintain the consistency of our estimates over time. This combination of methods enhances our understanding of the changing e-commerce landscape and urban freight patterns, informing strategic operations and policymaking in the urban delivery sector. We motivate the balance of the research by hypothesizing that last mile deliveries are mainly fulfilled by autonomous sidewalk robots (ASDRs) and propose energy-efficient GPS use in these robots by adjusting their data retrieval rate in response to new local conditions. This strategy, managed through a linear predictive filter, could be adapted for various delivery and location-based services.

5 - Assessing Fairness in the Presence of Missing and Mislabeled Sensitive Attributes

Hyungrok Do, NYU School of Medicine, New York, NY, United States

It is well-documented that sensitive attributes–which are often socio-economic status or demographic information–such as gender, race, and age, are frequently reported inaccurately in many databases. This is a significant challenge in the context of investigating predictive performance disparity since the metrics are calculated with the assumption that correct sensitive attribute information are provided. There is a relatively small amount of prior work in algorithmic fairness on handling of missing and mislabeled sensitive attributes. We propose a methodology to estimate per-patient probability distributions for sensitive attributes such as gender, race, and age, based on other known features of the patient. Using these estimated probability distributions, we calculate prediction disparity metrics, such as differences in accuracy or true positive rates between groups, robust to the missing and mislabeled sensitive attributes are. This will enable us to achieve robust estimates of fairness metrics and avoid misleading assessments of predictive disparities.

6 - Ensuring Fair Prediction with Group-Aware Priors

Tim G. J. Rudner, New York University, New York, NY, United States

We consider the problem of training neural network models that make fair predictions in settings where certain subpopulations, or groups, are highly underrepresented in the training data. To address this problem, we develop a family of group-aware prior distributions over neural network parameters explicitly designed to favor models that are fair. Priors in this family assign high probability density to parameter values that induce models that are robust across groups. We show how to construct group-aware priors and derive a tractable optimization objective that incorporates them into model training. We demonstrate empirically that training with group-aware priors yields fair models on several challenging prediction tasks. Our evaluation shows that group-aware priors outperform state-of-the-art methods on vision and language classification tasks while only requiring access to a small amount of group information. Group aware-priors are conceptually simple, complement alternative approaches for fair prediction, and allow harnessing Bayesian inference to further improve model robustness.

MD24

Summit - 344

Optimizing Digital Content, Revenue, and Education Strategies

Contributed Session

Chair: Amrita Thomas, University of North Texas, 3832 Greenhills Ct E, Denton, TX, 75038, United States

1 - Revenue Sharing Strategies among Digital Content Creators

Icheng Chiang, Fordham University, New York, NY, United States, Jhih-Hua Jhang-Li

Social media platforms foster creativity by helping creators monetize their content. During content creation, however, copyright disputes could arise when prior uploads were used as base materials for derivative work. One way to obtain copyright is to pay recurring royalty or a steep fee upfront—propositions that could be cost-prohibitive for many content creators. Recently, a major social media platform unveiled a revenue-sharing mechanism in which content creators split the ad revenue from their derivative works with copyright owners according to a predefined revenue-sharing ratio. We propose a mechanism to optimize the revenue-sharing ratio and to suggest when/whether the copyright owner should file the copyright claim. Upon receiving a claim, the content creator can cancel the upload, agree to revenue sharing, or dispute the claim within a time window. We show how our approach can complement flat-fee licensing and help reduce the likelihood of litigation.

2 - Managing the Digital Advertising Economy: Balancing Generated Content and Profitability

Na Liu, Cornell University, Ithaca, NY, United States, Chris Forman

This paper examines the influence of digital monetization strategies on user-generated content (UGC) platforms like YouTube and TikTok, which play a crucial role in facilitating collaborations between content creators and industrial sponsors for online media curation. We investigate the effects of commission fee deductions on various platform outcomes, including revenue sharing, total profits, ad selection, and users' activities. Through DID Models with ex-ante classification and robustness checks spanning over years, we find surprisingly that advertisers are shifting towards goods-selling links that experienced the fee deductions, resulting in the substitution of ad types. This transition concentrates the goods-selling market on specific popular products, bolstering direct sales on major online retailing sites. However, the platform does not witness a substantial increase in total profits despite decreased commission fees on certain ad links, prompting shifts in user behavior regarding ad types. On the creators' side, content quality declined on average for untreated groups, signified by a drop of 2 interactions per 100 viewers, while it remains the same for treated creators with deducted commission fees' ads. In summary, while the platform's design does not directly escalate profits, it channels advertisers towards specific ad types and goods, intensifying competition among creators and sustaining content quality, informing "smarter decision" for the UGC platforms. This paper is the first to empirically analyze the strategic implications of UGC platform commission fees at scale from a multi-sided perspective.

3 - Are Recruitment Postings Believable? A Theory-Driven Supervised Learning and Interpretable Analysis for Digital Fraudulent Recruitment Posting Behavior

Tianteng Wang, City University of Hong Kong, Hong Kong, Hong Kong, David Xu, Keng Siau

The number of recruitment postings on digital recruitment-hiring platforms has increased after the COVID-19 pandemic. However, the weak surveillance of these platforms, combined with the fact that most recruitment seekers have relatively low vigilance and strong desire for recruitment offers enables scammers to easily deceive recruitment seekers for their money and confidential information. The existing manual experience-based fraud recruitment posting detection approach is labor intensive, inefficient, and prone to misdetection. Therefore, how to efficiently detect the fraudulence of digital recruitment postings without human interventions becomes a major operational challenge. Guided by the design science approach, we are informed by a kernel theory of interpersonal deception theory and propose a novel text analytic framework to extract theoretically important and managerially meaningful constructs, such as completeness, association, clarity, and relevance, from textual recruitment postings as operationalized and contextualized features with textual statistics, lexicon-based methods, named entity recognizer, topic modeling, information entropy, sentiment analysis techniques, pre-trained large language model, and Generative Pre-trained Transformer (GPT) model to predict the fraudulence. In terms of evaluations, we empirically compare our designed framework with state-of-the-art general-purpose algorithms that can be adapted to tackle the fraud recruitment detection problem. Our extensive analysis demonstrates the efficiency and interpretability of our proposed framework in detecting the fraudulence of recruitment postings. Using an illustrative example, we further quantify the economic value of our theory-inspired design via a cost-revenue analysis. We also highlight the academic contributions and practical implications of our work to recruitment seekers, recruiters, and third-party recruitment-hiring platforms.

4 - Innovating Classrooms: Leveraging AI for Engaging Educational Experiences

Amrita Thomas, University of North Texas, Denton, TX, United States

In the evolving landscape of educational innovation, integrating digital technologies into academic frameworks stands at the frontier for advancement. This presentation explores the incorporation of digital tools within a university curriculum. By utilizing these tools to enhance creative processes such as brainstorming and media production, we highlight the synergistic potential between theoretical knowledge and practical digital application. We aim to offer a view on how we can amplify creativity and innovation and provides insights and strategies for embedding these technologies into educational practices to foster a new era of creative excellence.

5 - Membership Pricing Strategies for Digital Content Platforms: Balancing Native and Traditional Advertising

Wan Xiong, Shanghai University of Finance and Economics, ShangHai, China, People's Republic of, Lu Wang

We consider ad-supported media platforms with incomplete information about the disutility consumers' experience from exposure to advertising. We characterize the platform's optimal subscription fees and the level of two kinds of advertising respectively (i.e. traditional advertising and sponsored advertising, also known as native advertising), exploring whether the decision-making power regarding the level of native advertising should be ceded to content creators. Employing mechanism design and Kuhn-Tucker conditions enables us to systematically analyze the key factors influencing market outcomes. We demonstrate that the introduction of sponsored content advertising reduces the level of traditional advertising and increases the subscription fees. However, the platform's profit increases when the ratio of aversion to advertisements among type H consumers, as compared to type L consumers, is very high or moderate high. Additionally, we find that the platform should allow the creator to decide the level of sponsored ads under certain conditions.

MD25

Summit - 345

MathWorks/ChiAha

Invited Session

Technology Showcase

1 - Techno-economic Analysis with MATLAB: Microgrid Planning for Green Hydrogen Production

Chris Lee, MathWorks, Natick, MA, United States

Techno-economic analysis of microgrids is essential for assessing the financial viability and technical feasibility of integrating renewable energy sources. This analysis enables cost-effective strategies and technological innovations to enhance energy resilience, reduce carbon footprints, and accelerate the transition to sustainable energy.

In this session, we will present a live demonstration of how to develop a techno-economic analysis framework using optimization tools in MATLAB. We will cover the following:

- Modeling using the problem-based optimization workflow
- Solving using the linear programming solver
- Performing sensitivity and statistical analysis

As a case study, we will focus on planning a microgrid for green hydrogen production. We will explore how to determine the optimal power and energy ratings within a microgrid system incorporating renewable energy and energy storage. The goal is to power an electrolyzer to produce green hydrogen at the lowest cost and highest return on investment over a 20-year period.

2 - Introducing ChiAha – the Smart Manufacturing Digital Twin Toolkit

Andrew Siprelle, ChiAha, Tallassee, FL, United States

ChiAha can predict production line performance and OEE within 1% accuracy. High-fidelity modeling constructs with statistically modeled data-driven behavior. Provide answers to many of the questions related to the design, operation and improvement of lines for optimum OEE.

MD26

Summit - 346

Industry Job Search Panel

Panel Session

Job Placement Services

Co-Chair: Robin Lougee, KnitWell Group, Yorktown Heights, NY, United States

Co-Chair: Tom Fink, INFORMS, Catonsville, MD, United States

1 - Moderator Panelist

2 - Moderator Panelist

3 - Panelist

James Cochran, The University of Alabama, Tuscaloosa, AL, United States

4 - Panelist

Mei Zhang, OTIS, Lewisville, TX, United States

5 - Panelist

Rajeev Namboothiri, GE Vernova Research, Bangalore, India

6 - Panelist

Elizabeth Cubbage, Petco, Fort Collins, CO, United States

7 - Panelist

Vinod Cheriyan, Stitch Fix, Oakland, CA, United States

MD27

Summit - 347

AI Process Innovation: Connection to the Human Element

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Abhishek Deshmane, Georgia Institute of Technology, Atlanta, GA, United States

1 - The Uneven Impact of Generative AI on Entrepreneurial Performance

David Holtz, UC Berkeley, Berkeley, CA, United States, Nicholas Otis, Rowan Clarke, Solene Delecourt, Rembrand Koning

There is a growing belief that scalable and low-cost AI assistance can improve firm decision-making and economic performance. However, running a business involves a myriad of open-ended problems, making it hard to generalize from recent studies showing that generative AI improves performance on well-defined writing tasks. In our field experiment with 640 Kenyan entrepreneurs, we as- sessed the impact of AI-generated advice on small business revenues and profits. Participants were randomly assigned to a control group that received a standard business guide or to a treatment group that received a GPT-4-powered AI busi- ness mentor via WhatsApp. We are unable to reject the null hypothesis that generative AI access has no impact, but are able to rule out the large effect sizes reported by other studies of generative AI's economic impact. Our over- all null result masks treatment effect heterogeneity with respect to the baseline business performance of the entrepreneur: our point estimates suggest that high performers benefited by just over 15% from AI advice, whereas low performers did about 8% worse with AI assistance. Exploratory analysis of WhatsApp in- teraction logs shows that both groups sought the AI mentor's advice, but that low performers did worse because they sought help on more challenging business tasks. Our findings highlight the potential and limitations of generative AI to enable entrepreneurs across the globe.

2 - The Impact of AI-Enabled Technology on The Learning of The Gig Workers

Youngsoo Kim, University of Alabama, Tuscaloosa, AL, United States, Dmitry Mitrofanov, Yuqian Xu

This paper investigates the relationship between AI-enabled technology and the learning dynamics of gig workers, through a large-scale randomized field experiment partnered with a major US delivery platform. Specifically, we look into the balance between leveraging AI-enabled technology for immediate productivity gains and fostering a learning environment that ensures the sustainable development of gig workers' skills. We believe that understanding this interaction is important for platform operators to design user-supporting technologies and systems that consider both short-term productivity gains and long-term skill development in the gig economy.

3 - Can Machine Learning Target Health Care Fraud? Evidence from Medicare Hospitalizations

Shubhranshu Shekhar, Brandeis University, Waltham, MA, United States, Jetson Leder-Luis, Leman Akoglu

The US spends more than $4 trillion per year on health care, largely conducted by private providers and reimbursed by insurers. A major concern in this system is overbilling and fraud by hospitals, who face incentives to misreport their claims to receive higher payments. In this work, we develop novel machine learning tools to identify hospitals that overbill insurers, which can be used to guide investigations and auditing of suspicious hospitals for both public and private health insurance systems. Using large-scale claims data from Medicare, the US federal health insurance pro- gram for the elderly and disabled, we identify patterns consistent with fraud among inpatient hospitalizations. Our proposed approach for fraud detection is fully unsupervised, not relying on any labeled training data, and is explainable to end users, providing interpretations for which diagnosis, procedure, and billing codes lead to hospitals being labeled suspicious. Using newly collected data from the Department of Justice on hospitals facing anti-fraud lawsuits, and case studies of suspicious hospitals, we validate our approach and findings. Our method provides an 8-fold lift over random targeting of hospitals. We also perform a post-analysis to understand which hospital characteristics, not used for detection, are associated with suspiciousness.

4 - Enhancing Emergency Department Patient Flow with Artificial Intelligence

SeungJin Wang, Siena College, Loudonville, NY, United States, Ratna Chinnam

This study aims to optimize patient flow and reduce length of stay (LOS) in the emergency department (ED) through early task initiation. An artificial intelligence (AI) model is developed using real ED electronic health record data to predict LOS and admission decisions. A reinforcement learning (RL) approach determines the optimal timing to disseminate patient information to facilitate resource allocation. The RL agent leverages predictions on LOS and admission, as well as virtual ED environment data in simulations. By continuously learning the most effective broadcasting actions, the RL mechanism optimizes patient flow. Outcomes demonstrate timely patient updates through the AI-enabled orchestration can reduce wait times. This not only improves care delivery but also enhances healthcare provider productivity and satisfaction while minimizing costs. Future work includes extending early task initiation techniques to other hospital departments for comprehensive patient flow optimization.

MD28

Summit - 348

Leveraging Access to Data to Improve Access to Social Services

Invited Session

Service Science

Chair: Opher Baron, University of Toronto, Toronto, ON, Canada

Co-Chair: Shany Azaria, University of Toronto, Toronto, ON, Canada

1 - From Docket to Data: Unpacking Judicial Congestion Using Process- and Queue-Mining

Shany Azaria, University of Toronto, Toronto, ON, Canada

The court system, pivotal for social justice, is overflowed by congestion affecting social welfare, economic development, and access to justice. Despite its significance and operational complexity, being characterized by constrained resources, increasing demand, long in-process waiting time, impatient customers, and LOS measured in years, empirical research on court operations remains limited. This scarcity is largely attributable to challenges in accessing reliable and comprehensive data. This study employs NLP and AI tools to transform two decades of US Federal District Court case dockets into a detailed event log, leveraging process- and queue-mining to explore judicial congestion. By scraping, labeling, and analyzing millions of docket entries through operations management lenses, we are able to observe the case flow in the system and assess congestion impacts on case processing. Our findings illuminate the judicial workflow, offering insights for reducing congestion and enhancing system efficiency. This pioneering approach underscores the potential of data-driven analysis in court systems operations.

2 - Service Operations for Justice-on-TIME: a Data-Driven Queueing Approach

Ramandeep Randhawa, USC Marshall School of Business, Los Angeles, CA, United States, Nitin Bakshi, Jeunghyun Kim

We study judicial delays in the Supreme Court of India by applying concepts from service operations. Our goal is to elucidate the drivers of congestion, focusing on metrics such as the expected case-disposition time (delay) and the expected number of cases awaiting adjudication (pendency), and leverage this understanding to recommend operational interventions. We employ data-driven calibrated simulations to model the analytically intractable case-management queue. The life cycle of a case comprises two stages: pre-admission (before determining its merit for detailed hearings) and post-admission. Our methodology allows us to capture the queueing dynamics in which the judges are shared resources across the two stages. The sources of inefficiency that drive congestion include a misalignment between scheduling guidelines and judicial capacity, coupled with the requirement to schedule hearings in advance. Together, these factors inhibit utilization of shared capacity across the two-stage judicial queue. We demonstrate how interventions that account for these inefficiencies can successfully tackle judicial delay. In particular, scheduling to improve the allocation of time across pre- and post-admission cases can cut down the expected delay by as much as 65%.

3 - Modeling and Evaluating Policy Impacts on the United States Immigration Court System: An Application of Simulation and Data Science

Geri Dimas, Bryant University, Smithfield, RI, United States, Andrew Trapp, Renata Konrad

There is a significant and growing backlog in the United States immigration court system with over 2 million cases waiting to be heard. Due to large influxes of immigrants together with limited design and resources, the court system struggles to manage this growing backlog, resulting in delays that unnecessarily tax governmental and community resources. Leveraging data science and discrete event simulation we model the intricacies of this court system, deconstructing different elements and their respective complexity. We extend the baseline model to investigate alternative policies, seeking more efficient and equitable solutions to address the growing backlog.

4 - Neural Network Version of Mixed Effect Models

Mian Adnan, University of Northern Colorado, Greeley, CO, United States

The neural network structure for the mixed effect models has been developed.

5 - How Driving Automation Will Save Demand Responsive Transit

Antonio Antunes, Universidade de Coimbra, Coimbra Codex, Portugal, Amir Brudner, Anne Patricio, Gonçalo Gonçalves Duarte Santos, Moshe Ben-Akiva

This research investigates the viablility and potential advantages of deploying a Demand-Responsive Transit (DRT) service in low-demand areas, with a particular emphasis on how
driverless technology could influence this service. The analysis employs a three-step process.Inttially, it estimates the demand function based on service characteristic like fares and travel time. Subsequently, an integer linear programming (ILP) model is applied to optimize aspects such as fleet sized, vehicle routing, and the number or drivers/supervisors to achieve profit maximization. The final step involves assessing key performace indicators (KPIs) tp evaluate the service from multiple stakeholder perspectives. Findings reveal that although human-operated DRT services can improve consumer surplus, they are economically inefficient due to high operational costs and required subsidies, which can explain their frequent failure. In contrast, driverless DRT services emerge as a viable alternative, with substantially lower costs and improved social welfare
compared to human-operated DRT, especially when using low-capacity vehicles. Therefore,
policymakers and transit agencies should explore the potential of driverless technology as a means to provide a successful public transportation solution in rural and low-density areas where conventional fixed-route services fall short.

MD29

Summit - 420

ONR Program Review: Integer and Discrete Optimization

Invited Session

OPT: Integer and Discrete Optimization

Chair: David Phillips, Office of Naval Research, University of New Mexico, Catonsville, MD, United States

1 - Mixed-Integer Programming Approaches to Generalized Submodular Optimization and Its Applications

Simge Kucukyavuz, Northwestern University, Evanston, IL, United States

Submodularity is an important concept in integer and combinatorial optimization. A classical submodular set function models the utility of selecting homogenous items from a single ground set, and such selections can be represented by binary variables. In practice, many problem contexts involve choosing heterogenous items from more than one ground set or selecting multiple copies of homogenous items, which call for extensions of submodularity. We refer to the optimization problems associated with such generalized notions of submodularity, Generalized Submodular Optimization (GSO). GSO is found in wide-ranging applications, including infrastructure design, healthcare, online marketing, and machine learning. Due to the often highly nonlinear (even non-convex and non-concave) objective function and the mixed-integer decision space, GSO is a broad subclass of challenging mixed-integer nonlinear programming problems. In this tutorial, we first provide an overview of classical submodularity. Then we introduce two subclasses of GSO, for which we propose polyhedral theory for the mixed-integer set structures that arise from these problem classes. Our theoretical results lead to efficient and versatile exact solution methods that demonstrate their effectiveness in practical problems using real-world datasets.

2 - Dyadic Linear Programming

Gerard Cornuejols, Carnegie Mellon University, Pittsburgh, PA, United States, Ahmad Abdi, Bertrand Guenin, Levent Tuncel

A rational number is *dyadic* if it has a finite binary representation. Dyadic rationals are important for numerical computations because they have an exact representation in floating-point arithmetic on a computer. A vector is *dyadic* if all its entries are dyadic rationals. We study the problem of finding a dyadic optimal solution to a linear program, if one exists. We show how to solve dyadic linear programs in polynomial time. We give bounds on the size of the support of a solution as well as on the size of the denominators. To illustrate this theory, we present recent computational results obtained in collaboration with PhD students Anthony Karahalios and Vrishabh Patil. In a different direction, we present results on directed graphs obtained in collaboration with PhD students Olha Silina and Siyue Liu. The work of Gerard Cornuejols, Vrishabh Patil, Siyue Liu and Olha Silina is supported by ONR grant N00014-22-1-2528.

3 - Sensitivity Analysis for Mixed Binary Quadratic Programs

Santanu Dey, ISyE Georgia Tech, Atlanta, GA, United States, Diego Cifuentes, jingye xu

In many applications it is necessary to solve challenging Mixed Binary Quadratic Programming (MBQP) instances repeatedly. These instances typically have the same constraint matrix, representing data corresponding to some invariant physical infrastructure, but varying objective function and right-hand-side (rhs). This motivates the need to develop methodology for conducting sensitivity analysis of MBQPs. In this talk, we consider sensitivity analysis for MBQPs with respect to changing rhs. We show that even if the optimal solution of a given MBQP is known, it is NP-hard to approximate the change in objective function value with respect to changes in rhs. Next, we study algorithmic approaches to obtaining dual bounds for MBQP with changing rhs. We leverage Burer’s completely-positive (CPP) reformulation of MBQPs. Its dual is an instance of co-positive programming (COP) and can be used to obtain sensitivity bounds. We prove that strong duality between the CPP and COP problems holds if the feasible region is bounded or if the objective function is convex, while the duality gap can be strictly positive if neither condition is met. We also show that the COP dual has multiple optimal solutions, and the choice of the dual solution affects the quality of the bounds with rhs changes. Finally, we provide an algorithmic approach to find ``best values" of optimal dual solutions, and present preliminary computational results on sensitivity analysis for MBQPs.

4 - Representing Integer Program Value Function with Neural Networks

Andrew Schaefer, Rice University, Houston, TX, United States, Tu Nguyen, Joey Huchette

We study the value function of an integer program (IP), which characterizes its optimal objective as a function of its right-hand side. We show that the IP value function can be approximated to any desired degree of accuracy using neural networks (NNs). We do this by deriving a connection between Chvatal-Gomory (CG) cuts and the IP value function, resulting in a novel NN architecture that interprets the weights of the NN as CG multipliers. Using this NN architecture, we propose a new unsupervised training method specialized for learning the IP value function. Preliminary computational results show that our training method on data without labels is comparable with supervised training.

MD30

Summit - 421

Optimization Software Session II

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelmann, Arizona State University, Tempe, AZ, United States

1 - What's New in Fico Xpress Solver

Michael Perregaard, FICO, Birmingham, United Kingdom, Alexander Biele

We will give an overview of the latest enhancements, the newest features, and the most recent performance improvements in the FICO Xpress Solver for mixed-integer linear and nonlinear optimization problems. These include new heuristics, cutting and branching techniques, a more streamlined API, and updates to our global MINLP solver.

2 - Latest Developments in the Artelys Knitro Optimization Solver

Richard Waltz, Artelys, Los Angeles, CA, United States

Artelys Knitro is the premier solver for nonlinear optimization problems. Knitro offers both interior-point and active-set algorithms for continuous models, as well as tools for handling problems with integer variables and other discrete structure. This talk will highlight the latest developments in Knitro, focusing on some of the recent advances in solving mixed-integer nonlinear problems (both convex and non-convex), and heuristics for global optimization. We will also present Knitro results on large-scale models demonstrating speedups achieved with the latest Knitro releases.

3 - Latest Progress in Optimization Software

Hans Mittelmann, Arizona State University, Tempe, AZ, United States

An overview will be given on the performance of the latest optimization software based on our benchmarks

4 - Recent Developments in Optimization with MATLAB

Steve Grikschat, MathWorks, Natick, MA, United States

We will present an overview of the latest developments in MATLAB optimization. MATLAB has solvers for engineering and scientific applications where problems range from linear to nonlinear, mixed-integer, and non-smooth.

MD31

Summit - 422

Learning and Applications of Mean Field Models and Multi Agent Systems

Invited Session

OPT: Machine Learning

Chair: Gokce Dayanikli, University of Illinois Urbana-Champaign, Champaign, IL, United States

Co-Chair: Mathieu Lauriere, New York University Shanghai, Shanghai, N/A

1 - Synchronization Games

H. Mete Soner, Princeton University, Princeton, NJ, United States

Building on Winfree's work, the Kuramoto model (1975) has become the corner stone of mathematical models of collective synchronization, and has received attention in all natural sciences, engineering, and mathematics. While the classical model postulates the dynamics of each oscillator in the form of a system of nonlinear ordinary differential equations, Yin, Mehta, Meyn, & Shanbhag (2010) use the mean-field game (MFG) formalism of Lasry & Lions, and Huang, Caines, & Malhame. In this talk, in addition to the Yin et.al model, we also introduce a simpler two state model which can be seen as a discretization of the original one. We outline results showing that the mean field approach also delivers same type of results including the phase transition from incoherence to synchronization. In particular, in the discrete setting we provide a comprehensive characterization of stationary and dynamic equilibria along with their stability properties. In all models, while the system is unsynchronized when the coupling is not sufficiently strong, fascinatingly, they exhibit an abrupt transition to a full synchronization above a critical value of the interaction parameter. In the subcritical regime, the uniform distribution representing incoherence is the only stationary equilibrium. Above the critical interaction threshold, the uniform equilibrium becomes unstable and there is a multiplicity of stationary equilibria that are self-organizing. The discrete model with discounted cost present dynamic equilibria that spiral around the uniform distribution before converging to the self-organizing equilibria. With an ergodic cost, however, unexpected periodic equilibria around the uniform distribution emerge.

2 - Stochastic differential games on graphs

Haosheng Zhou, University of California, Santa Barbara, Santa Barbara, CA, United States, Ruimeng Hu, Jihao Long

In this talk, we present our recent study unveiling the connection between graph properties and stochastic differential games. We begin by introducing a game modeled on graphs and derive the Markovian Nash equilibrium via Riccati equations. We then analyze the iterative scheme – fictitious play – for solving the Riccati equations and demonstrate its convergence rate, which is highly correlated with the underlying graph structure. Additionally, we present a semi-explicit expression for the Markovian Nash equilibrium in games played on vertex-transitive graphs. This expression serves as a benchmark to test our proposed deep learning algorithms.

3 - Partially-observable and major minor mean field control for multi-agent reinforcement learning

Kai Cui, Technische Universität Darmstadt, Darmstadt, Germany

In this talk, we consider mean field control in discrete time with partial observability and major agents for multi-agent reinforcement learning. Here, partial observability refers to limited information available to agents under a deterministic limiting mean field, whereas major agents refer to additional arbitrary, general agents under limiting stochastic mean fields. We begin by presenting standard mean field control and associated learning algorithms. Then, we relax the two limitations of full observability and weak interaction via novel mean field control models. We also give associated multi-agent reinforcement learning policy gradient algorithms and analyze the error in the policy gradient estimation. Finally, we briefly give an outlook on possible future extensions and current limitations.

4 - Stochastic filtering equations for diffusions on infinite graphs

Ichiba Tomoyuki, University of California-Santa Barbara, Santa Barbara, CA, United States

We discuss stochastic filtering problem for diffusions on infinite graphs and its applications to machine learning. A prototypical example we consider here is the problem for diffusions which are described by the stochastic differential equations of directed chain interaction and mean-field interactions on the infinite directed graph. We derive the linear and non-linear systems of the corresponding filtering equations and examine its solvability and properties which include the smoothness of the solution. In this talk we introduce its applications to the generative adversarial network problem.

MD32

Summit - 423

Advances of Reformulation Methods in General Integer Programming

Invited Session

OPT: Global Optimization

Chair: Ningji Wei, Texas Tech University, Lubbock, TX, United States

1 - Rank-One Convexification for Quadratic Functions with Sign-Controlled Variables

Soobin Choi, University of Southern California, Los Angeles, CA, United States, Valentina Cepeda, Andres Gomez, Shaoning Han

In this talk, we study the epigraph of a quadratic form, whose variables are ``sign-controlled'' by indicator variables. We describe the convex-hull for a special case of the set, in which the quadratic term is defined by a rank-one matrix. From the rank-one convexification, we derive a convex relaxation for general quadratic functions defined by positive semi-definite matrices.

2 - Binary Integer Program Reformulation: a Set System Approximation Approach

Ningji Wei, Texas Tech University, Lubbock, TX, United States

We present a generic reformulation framework for binary integer programs (BIPs) without imposing additional specifications for the objective function or constraints. To facilitate such generality, we introduce a set system approximation theory designed to identify the tightest inner and outer approximations for any binary solution space using special types of set systems. This development leads to an exact reformulation framework for general BIPs, centered around set covering and subtour elimination inequalities. We investigate the implications of this methodology on various instance problems, uncovering new solution strategies and structural insights. Building upon these advancements, we also extend the classic max-flow min-cut theorem into a broader context of set system duality. Overall, our framework explores a new direction in the field of integer programming by examining the algebraic properties of set systems and their operators, which may spur additional research questions and further enrich the field.

3 - A New Branching Rule for Optimizing Range and Other Order-Based Objective Functions

Rui Chen, Cornell Tech, Cornell University, New York, NY, United States, Bart van Rossum, Andrea Lodi

We consider range minimization problems featuring exponentially many variables, as frequently arising in fairness-oriented or bi-objective optimization. While branch-and-price is successful at solving cost-oriented problems with many variables, the performance of classical branch-and-price algorithms for range minimization is drastically impaired by weak linear programming relaxations. We propose range branching, a generic branching rule that directly tackles this issue and is compatible with any problem-specific branching scheme. We show several desirable properties of range branching and show its effectiveness on a series of fair capacitated vehicle routing instances. Range branching significantly improves multiple classical branching schemes in terms of computing time, optimality gap, and size of the branch-and-bound tree, allowing us to solve many more large instances than classical methods. We also consider a generalization of range branching to deal with more general order-based objective functions such as Gini deviation.

4 - Optimal Piecewise Quadratic Approximations for Enhancing Deterministic Global Optimization

Yingkai Song, Princeton University, Princeton, NJ, United States, Christos Maravelias

A range of process systems engineering problems can be formulated as mathematical optimization models, though the resulting, often nonlinear, models are computationally challenging. A common solution strategy involves simplifying nonlinearities through piecewise linear (PWL) approximations. However, due to the linearity of each approximation segment, this method may require many segments to achieve a desired accuracy, subsequently increasing the number of necessary binaries and thus complexity. Given recent advances in numerical solvers such as Gurobi, which can solve mixed-integer quadratically constrained programs (MIQCP) efficiently, there's a motivation to use piecewise quadratic (PWQ) approximations instead. Employing PWQ approximations enables using significantly fewer segments, which has the potential to outperform models/solvers using PWL approximations.

This presentation introduces new approaches for constructing optimal PWQ approximations with both the placement of break points and the approximation coefficients being decision variables. The objective is either to minimize approximation error given a maximum number of segments or minimize the number of segments given a maximum approximation error. Our approach builds upon existing studies for PWL approximation and introduces mixed-integer models to construct optimal PWQ approximations for two-dimensional discrete data. A sampling-and-refining strategy subsequently extends these models to approximate univariate scalar functions. Numerical examples show how using the proposed PWQ approximation can lead to improved computational performance compared to using state-of-the-art PWL approximations or the original models.

MD33

Summit - 424

Distributional Robustness and Sequential Decision-Making

Invited Session

OPT: Optimization Under Uncertainty

Chair: Mohammed Amine Bennouna, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Multistage Distributional Robustness, TIME Consistency and Optimal Policy

Rui Gao, University of Texas at Austin, Austin, TX, United States

In this talk, I will discuss distributional robustness in a dynamic setting, where the deviation is measured by the causal transport distance between stochastic processes. Such a choice accounts for information evolution, making it hedge against a plausible family of data processes. I will present a recursive reformulation to evaluate the worst-case risk of any given random sequence and explore the intricacies of time consistency in dynamic risk measures. Furthermore, I will present dynamic programming reformulations for finding the optimal robust policy.

2 - The Curious Price of Distributional Robustness in Reinforcement Learning with a Generative Model

Yuting Wei, University of Pennsylvania, PHILADELPHIA, PA, United States

In this talk, we investigate model robustness in reinforcement learning (RL) to reduce the sim-to-real gap in practice. We adopt the framework of distributionally robust Markov decision processes (RMDPs), aimed at learning a policy that optimizes the worst-case performance when the deployed environment falls within a prescribed uncertainty set around the nominal MDP. Despite recent efforts, the sample complexity of RMDPs remained mostly unsettled regardless of the uncertainty set in use. It was unclear if distributional robustness bears any statistical consequences when benchmarked against standard RL. Assuming access to a generative model that draws samples based on the nominal MDP, we characterize the sample complexity of RMDPs when the uncertainty set is specified via either the total variation (TV) distance or χ^2 divergence.

3 - Sample Complexity of Inventory Control with Fixed Ordering Cost

Xiaoyu Fan, Stern School of Business, New York University, New York, NY, United States, Zhengyuan Zhou

We show in this work that a class of structured MDPs admits more efficient learning (i.e., lower sample complexity bounds) compared to the best possible/known algorithms in generic RL. We focus on the MDPs describing the inventory control system with fixed ordering costs, a fundamental problem in supply chains. We develop an algorithm applied to the inventory MDPs, which leads to strictly lower sample complexity bounds compared to the optimal or best-known bounds recently obtained for the general MDPs. We improve on those ``best-possible'' bounds by carefully leveraging the structural properties of the inventory dynamics in various settings.

4 - Statistical Learning of Distributionally Robust Stochastic Control in Continuous State Spaces

Shengbo Wang, Stanford University, Stanford, CA, United States, NIAN SI, Jose Blanchet, Zhengyuan Zhou

We explore the control of stochastic systems with potentially continuous state and action spaces, characterized by the state dynamics $X_{t+1} = f(X_t, A_t, W_t)$. Here, X, A, and W represent the state, action, and exogenous random noise processes, respectively, with f denoting a known function that describes state transitions. Traditionally, the noise process $\{W_t, t >= 0\}$ is assumed to be independent and identically distributed (i.i.d.), with a distribution that is either fully known or can be consistently estimated. However, the occurrence of distributional shifts, typical in engineering settings, necessitates the consideration of the robustness of the policy. This paper introduces a distributionally robust stochastic control paradigm that accommodates possibly adaptive adversarial perturbation to the noise distribution within a prescribed ambiguity set. We examine two adversary models: current-action-aware and current-action-agnostic, leading to different dynamic programming equations. Furthermore, we characterize the optimal finite sample minimax rates for achieving uniform learning of the robust value function across continuum states under both adversary types, considering ambiguity sets defined by f_k-divergence and Wasserstein distance. Finally, we demonstrate the applicability of our framework across various real-world settings.

MD34

Summit - 425

Game Theory and Optimization: Minimax and Variational Inequality Approaches

Invited Session

OPT: Nonlinear Optimization

Chair: Afrooz Jalilzadeh, The University of Arizona, Tucson, United States

1 - Variance Reduced Halpern Iteration for Finite-Sum Monotone Inclusions

Xufeng Cai, University of Wisconsin-Madison, Madison, WI, United States, Ahmet Alacaoglu, Jelena Diakonikolas

Machine learning approaches relying on such criteria as adversarial robustness or multi-agent settings have raised the need for solving game-theoretic equilibrium problems. Of particular relevance to these applications are methods targeting finite-sum structure, which generically arises in empirical variants of learning problems in these contexts. Further, methods with computable approximation errors are highly desirable, as they provide verifiable exit criteria. Motivated by these applications, we study finite-sum monotone inclusion problems, which model broad classes of equilibrium problems. Our main contributions are variants of the classical Halpern iteration that employ variance reduction to obtain improved complexity guarantees in which n component operators in the finite sum are “on average” either cocoercive or Lipschitz continuous and monotone, with parameter L. The resulting oracle complexity of our methods, which provide guarantees for the last iterate and for a (computable) operator norm residual, is $\tilde O(n + \sqrt{n}L \epsilon^{-1})$, which improves upon existing methods by a factor up to $\sqrt{n}$. This constitutes the first variance reduction-type result for general finite-sum monotone inclusions and for more specific problems such as convex-concave optimization when operator norm residual is the optimality measure. We further argue that, up to polylogarithmic factors, this complexity is unimprovable in the monotone Lipschitz setting; i.e., the provided result is near-optimal.

2 - Projection-free Methods for Solving Nonconvex-concave Saddle Point Problems

Morteza Boroun, The University of Arizona, Tucson, AZ, United States, Erfan Yazdandoost Hamedani, Afrooz Jalilzadeh

This study addresses constrained saddle point (SP) problems with nonconvex-(strongly) concave and smooth objective functions, which are widely applicable in problems arising in machine learning such as robust multi-class classification and dictionary learning. While projected gradient-based primal-dual methods exist, those with projection-free oracles are limited. In this work, we propose new projection-free primal-dual algorithms with convergence rate guarantees in both deterministic and stochastic scenarios.

3 - Linear Convergence Rate for Quadratic Functional Growth Saddle Point Problems

Cody Melcher, The University of Arizona, Tucson, AZ, United States, Erfan Yazdandoost Hamedani, Afrooz Jalilzadeh, Conner Alubowicz

In this talk, we consider a convex-concave saddle point problems and how to incorporate quadratic functional growth assumptions into them. Saddle point problems arise in a variety of areas such as machine learning, signal processing, and Nash equilibrium. While it has been shown that strongly convex-strongly concave setting leads to a linear convergence guarantee, we consider a weaker assumption based on the quadratic functional growth to ensure the same guarantee. Finally, we present numerical experiments to validate theory.

4 - First-order Methods for Stochastic Variational Inequality Problems with Function Constraints

Mohammad Khalafi, Southern methodist university, DALLAS, TX, United States, Digvijay Boob

The monotone Variational Inequality (VI) is a general model with important applications in various engineering and scientific domains. In numerous instances, the VI problems are accompanied by function constraints that can be data-driven, making the usual projection operator challenging to compute. This paper presents novel first-order methods for the function-constrained VI (FCVI) problem under various settings, including smooth or nonsmooth problems with stochastic operators and/or stochastic constraints. First, we introduce the AdOpEx method, which employs an operator extrapolation scheme of the KKT operator of the FCVI in a smooth deterministic setting. This operator is not uniformly Lipschitz continuous in the Lagrange multipliers. We show that an adaptive selection of stepsizes leads to bounded multipliers and achieves the optimal O(1/T) convergence rate. A direct application of such techniques to nonsmooth and stochastic FCVIs is challenging. Hence, we introduce design changes to the AdOpEx method and propose a novel P-OpEx method that takes a partial extrapolation. We show a convergence rate of $O(1/\sqrt{T})$ for P-OpEx when both the operator and constraints are stochastic. This method has suboptimal dependence on the noise and Lipschitz constants of function constraints. We propose a constraint extrapolation type approach leading to the OpConEx method that improves this dependence by an order of magnitude. Moreover, all our algorithms can easily extend to saddle point problems with function constraints that couple the primal and dual variables. For such problems, our algorithms maintain the same complexity results for the aforementioned cases.

MD35

Summit - 427

TSL Dissertation Award and Information for the TSL Data-Driven Research Challenge

Award Session

Transportation Science and Logistics (TSL)

Chair: Margaretha Gansterer, University of Klagenfurt, Klagenfurt, Austria

Co-Chair: Lei Zhao, Tsinghua University, Beijing, N/A

Co-Chair: Hai Wang, Singapore Management University, Singapore, Singapore

1 - Traffic Signal Optimization with Connected Vehicle Trajectories

Xingmin Wang, University of Michigan, Ann Arbor, MI, United States

Traffic light optimization is known to be a cost-effective method for reducing congestion and energy consumption in urban areas without changing physical road infrastructure. However, due to the high installation and maintenance costs of vehicle detectors, most intersections are controlled by fixed-time traffic signals that are not regularly optimized. To alleviate traffic congestion at intersections, we present a large-scale traffic signal re-timing system that uses a small percentage of vehicle trajectories as the only input without reliance on any detectors. We develop the probabilistic time-space diagram, which establishes the connection between a stochastic point-queue model and vehicle trajectories under the proposed Newellian coordinates. This model enables us to reconstruct the recurrent spatial-temporal traffic state by aggregating sufficient historical data. Optimization algorithms are then developed to update traffic signal parameters for intersections with optimality gaps. A real-world citywide test of the system was conducted in Birmingham, Michigan, and demonstrated that it decreased the delay and number of stops at signalized intersections by up to 20% and 30%, respectively. This system provides a scalable, sustainable, and efficient solution to traffic light optimization and can potentially be applied to every fixed-time signalized intersection.

2 - TSL information on the data-driven research challenge

Hai Wang, Singapore Management University, Singapore, Singapore, Lei Zhao

TSL and Meituan, a platform offering location-based consumer products and services including entertainment, dining, delivery, travel, and more, are partnering to provide researchers and practitioners worldwide with access to Meituan’s operational-level food delivery data in the first INFORMS TSL Data-Driven Research Challenge. This initiative aims to foster data-driven research on real-world challenges. Participants are expected to develop data-driven optimization models using the real data provided by Meituan to address some of the suggested questions or to explore questions of their own interest.

In this information session, we aim to cover the following aspects: (1) the academic and industry background of the research challenge; (2) the general administration of the challenge; (3) some descriptive analysis results and insights from the data provided; and (4) answer any questions from the attendees.

3 - Information About the TSL Data-Driven Challenge

Lei Zhao, Tsinghua University, Beijing

MD36

Summit - 428

Behavioral-informed Analytics and Decision-making for Urban Transportation Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Tianming Liu, University of Michigan, Ann Arbor, United States

1 - On the Economic Value of Electric Vehicle Charging Infrastructure

Daniel Vignon, New York University, New York, NY, United States, Yanchao Li, Prakrith Raja

In 2021, the Infrastructure Investment and Jobs Act (IIJA) allocated $7.5 bn for the construction of a national network of electric vehicle (EV) chargers. These are to be distributed to States according to a set of criteria which includes: "(…) meeting current and anticipated market demands for electric vehicle charging infrastructure, including with regard to power levels and charging speed, and minimizing the time to charge current and anticipated vehicles". Satisfying such a criterion naturally involves tradeoffs which must balance costs and benefits of a given charging deployment strategy. However, evaluating these tradeoffs when it comes to EV charging infrastructure is non-trivial and involves accounting for the interactions between multiple forces: activity scheduling, parking competition, and adoption dynamics. Our proposed research seeks to tackle that challenge. We propose and estimate a dynamic equilibrium framework that allows us, through counterfactuals in the context of New York City, to assess the economic value of charging infrastructure. Our insights shed light not only on behavioral patterns that may help--or hinder--EV adoption but also on efficient infrastructure policy.

2 - Principled Data Fusion to Infer Sociodemographic Variables from Big Mobility Data: a Semi-Supervised Bayesian Gp-Lvm Approach

Ekin Ugurel, University of Washington, Seattle, WA, United States, Khoa Dang, Prateek Bansal, Cynthia Chen

The widespread use of mobile devices and the omnipresence of network connectivity have led to an enormous volume of timestamped spatial data. While longitudinal, this passively collected mobile (PCM) data lacks the rich contextual insights of the more traditional household travel survey (HTS). We propose a principled method to impute latent factors (such as sociodemographic variables) into GPS-based datasets for a variety of use cases, like enhancing the performance of predictive models or enabling generalizable travel behavior inferences over time. Our method is both behaviorally- and spatially-informed—for the former, it uses a set of "labeled" trajectories (i.e., ones with the latent factors we seek to impute) to learn the relationship between travel behavior and various sociodemographic variables, whereas for the latter, it considers the underlying census tract at which the individual resides. Additionally, the word "principled" refers to our model's handling of uncertainty—rather than predicting point estimates, we propagate the uncertainty inherent in our learning of latent factors forward. We achieve this with a Bayesian treatment of the Gaussian Process latent variable model (GP-LVM), which allows us to report confidence bounds in any of our use cases. We demonstrate the robustness of our approach in two case studies.

3 - Cyberattack Impacts in Mixed-Flow Traffic: Driving Simulator Study

Yangjiao Chen, Georgia Institute of Technology, Atlanta, GA, United States, Shubham Agrawal, Srinivas Peeta

Connected and autonomous vehicles (CAVs) are expected to generate safety and efficiency benefits for the transportation system. However, CAVs are vulnerable to cyberattacks, potentially creating mixed-flow traffic of human-driven vehicles (HDVs), normal CAVs and compromised CAVs. To understand the associated network-level impacts, there is a need to analyze the interactions between HDVs and compromised CAVs and how they differ from the interactions between HDVs and normal CAVs. In this context, a driving simulator-based study is designed to investigate HDV car-following (CF) behavior in mixed-flow traffic when the predecessor CAV is normal or compromised. The effects of cybersecurity awareness and perception on CF behavior are considered. Statistical analysis and CF model calibration are performed based on the trajectory data collected from the driving simulator experiments to examine impacts on string stability and safety. The results also provide insights for CF behavior prediction under cyberattacks and mitigation measures to enhance CAV cybersecurity to assure putative traffic flow benefits.

MD37

Summit - 429

Innovative Last-Mile Solutions for Passengers and Freight

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Zhengtian Xu, The George Washington University, Washington, DC, United States

Co-Chair: Kenan Zhang, EPFL, Lausanne, Switzerland

1 - Optimizing TIME of Arrival Rather than Estimation: Joint Operation with Pricing for on-Demand Food Delivery Systems

Kaihang Zhang, The University of Hong Kong, Hong Kong, China, People's Republic of, Jintao Ke

The on-demand food delivery (OFD) system relies highly on time-sensitive operations and well-designed pricing policies. Therefore, these two aspects are crucial factors in the decision-making process for OFD operations. Optimizing pricing policy is a common practice in on-demand service operations including OFD as it influences the delivery supply-demand relationship and ultimately the platform's profitability or other metrics. On the other hand, the estimated time of arrival (ETA) is not typically regarded as a decision variable but rather as a prediction. However, it is the direct implementation of time sensitivity because it tells customers when their food will arrive and also shows delivery deadlines to drivers. As a result, the change in ETA decision may impact both customer demand and drivers' willingness to deliver. This paper considers the ETA as one of the platform's decision variables together with pricing. We present a joint optimization model of OFD which uses the unique characteristics of the OFD market, i.e., the order bundling behavior of drivers, as a medium to analyze the supply and demand equilibrium of delivery requests. We present insights into the operations of the OFD market in terms of ETA and pricing using the proposed mathematical model and show the non-monotonic impact of both decision factors.

2 - Analysis of Microhubs for Three-Sided Meal Delivery Services

Linxuan Shi, The George Washington University, Washington, DC, United States, Zhengtian Xu

In recent years, online food ordering and delivery platforms, such as DoorDash, Uber Eats, and Meituan, have experienced rapid growth in their three-sided market form. These platforms connect a range of restaurant suppliers to the people in surrounding neighborhoods and enable efficient meal delivery through a network of crowdsourced deliverers. This study examines the operational mechanism of microhubs in meal delivery services, acting as transshipment points where meal packages can be temporarily stored and consolidated for batch delivery. We develop aggregate analytical models for delivery operations with and without microhubs and compare their relative performances under various market conditions. Critical system metrics, with a specific focus on customers' and deliverers' experiences, are formulated to analyze and evaluate the potential efficiencies achievable through microhubs.

3 - The Parcel Delivery Problem Considering Crowd-Shipping and Uncertainty of Customer Demands

Sisi Jian, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Shixuan Tang, Wentao Huang

Express companies are faced with lower economic efficiency if they continue offering direct delivery service to customers because of the expensive fulfillment cost of last-mile delivery and the rapid growth of parcel volume. This paper therefore introduces an emerging logistic service, crowd-shipping, to the problem. Packages destined for remote areas can be dropped at predetermined service points to avoid costly delivery expense, and crowd-shipping drivers in the nearby regions will be designated as relay couriers from the service points to customers. A mixed-integer linear programming (MILP) model is then formulated with the crowd-shipping technique to determine decisions on the fleet size, the crowd-shipping service, and the routing choice. Due to the computational complexity of the problem, a heuristic algorithm based on the adaptive large neighborhood search (ALNS) is tailored to solve large-scale instances. Numerical experiments are conducted in comparison with other solution means and a deterministic model to validate the effectiveness and efficiency of the proposed method. This paper is expected to offer an economic delivery approach for express companies.

4 - On-Demand Food Delivery with Compensation Services

Yanling Zhuang, McMaster University, Hamilton, ON, Canada, Elkafi Hassini, Yun Zhou, Yufei Yuan

In food delivery services, on-time delivery is critical because customers base their orders on estimated arrival times and delivery fees. Failure to meet these promises results in delivery delays and potential loss of future orders. To counteract these issues, platforms may offer compensation to customers for delays. This paper presents an analytical model that examines how promised delivery times affect the decisions of platforms, customers, and drivers, and the impact of customer compensation policies.

MD38

Summit - 430

Airfield and Airspace Management

Invited Session

Aviation Applications

Chair: Wayne Ng, Singapore University of Technology and Design, Singapore, Singapore

1 - Modeling Demand Competition in Long-haul Aviation Markets

Nuno Ribeiro, Singapore University of Technology and Design, Singapore, Singapore, Mingmei Li, Sebastian Birolini

We look at the strategic hub competition problem in long-haul markets at a global scale. We identify itinerary attributes affecting passengers' choices and integrate them into demand models combined with optimization to analyse competition across airport hubs.

2 - Predicting Convective Weather in the Terminal Airspace of Singapore

Jetanat Datephanyawat, Singapore University of Technology and Design, Singapore, Singapore, Nuno Ribeiro

We develop probabilistic models to predict lightning occurrences in the terminal airspace of Singapore. Leveraging geostationary data, including cloud contours and lightning occurrences, we construct a suite of spatial tracking features to enhance the prediction of convective weather movements. Subsequently, we design algorithms to optimize aircraft trajectories, incorporating the calculated lightning risk probabilities to ensure safer and more efficient routing.

3 - Optimizing Aircraft Flows and Fuel Savings in Terminal Airspace Operations

Wayne Ng, Singapore University of Technology and Design, Singapore, Singapore, Sebastian Birolini, Nuno Ribeiro

In this research, we aim to optimize aircraft flows in the terminal airspace, also taking into account environmental considerations through minimizing fuel consumption, at a strategic level. To optimize aircraft flows, we previously developed an integrated optimization approach that utilizes matheuristic algorithms to optimize runway aircraft sequencing and decisions in the terminal airspace, including aircraft speeds, utilization of holding stacks, and vectoring. This model combines a Linear Programming model with metaheuristics, such as a Genetic Algorithm, to give near-optimal solutions in reasonable computation times. In order to also optimize fuel consumption, we propose 2 models, i) an extension of the LP to also consider altitude and speeds and ii) a data-driven approach to estimate fuel consumption between waypoint-pairs, which can also account for uncertainty. We utilize one year of high quality satellite data of flights in the congested terminal airspace of Changi Airport in Singapore, to calibrate both the parameters for the first model and the data-driven fuel estimation in the second model. Using a real flight schedule of Changi Airport, we test and discuss the applicability of our models. In this study, we also investigate the potential trade-offs between flight delays and total fuel consumption.

4 - Vehicle Routing Models for Efficient Airside Operations

Marcus Koh, Singapore University of Technology and Design, Singapore, Singapore, Peter Jackson, Nuno Ribeiro

In this research, we aim to develop a mesoscopic simulation tool to analyse airside operations and optimise resource planning for related operations. As a starting point, we focus on the development of vehicle routing programs to model the aircraft turnaround operations for purposes of resource optimization.

MD39

Summit - 431

RAS Round Table on Railway Service Reliability - Part 2

Panel Session

Railway Applications

Co-Chair: Erick Wikum, Wikalytics LLC, 4906 Whispering Creek Ct, Maineville, OH, 45039, United States

Co-Chair: Michael Gorman, University of Dayton, 300 college park, Dayton, OH, 45419, United States

1 - Panelist

Michael Gorman, University of Dayton, Dayton, OH, United States

2 - Panelist

Carl Van Dyke, TransNetOpt, West Windsor, NJ, United States

3 - Panelist

Edwin Kraft, TEMS, Inc., Frederick, MD, United States

4 - Panelist

Stefano Rieppi, Norfolk Southern Railway, Roswell, GA, United States

5 - Panelist

Jeremiah Dirnberger, Amtrak, Jacksonville, FL, United States

6 - Panelist

Sanford Sexhus, Oliver Wyman, Dallas, TX, United States

7 - Moderator Panelist

Erick Wikum, Wikalytics LLC, Maineville, OH, United States

MD40

Summit - 432

Emerging Topics in Behavioral Operations

Invited Session

Behavioral Operations Management

Chair: Basak Kalkanci, Georgia Tech, Atlanta, GA, United States

Co-Chair: Swanand Kulkarni, Tuck School of Business, Dartmouth College, Hanover, NH, United States

1 - Trade Credits and Visit Frequency: the Role of Order Financing on Nanostore Logistics Efficiency

Rafael Escamilla, Arizona State University, Tempe, AZ, United States, Jan Fransoo, Marcos Mogollon

Millions of mom-and-pop nanostores dominate the grocery retail landscape in emerging markets. As nanostores are cash and storage space constrained, their suppliers tend to visit them with a high frequency, causing high operational costs. It is unclear if credit could mitigate such costs by allowing for a lower visit frequency. To investigate this, we conduct a randomized field experiment on nanostores that have not made recent use of supplier credit, with the objective to uncover how trade credits interact with the visit frequency and with the available storage space to shape shopkeepers' ordering behavior. We find that trade credits moderate the positive relationship between visit frequency and bi-weekly order size, with credit effects being salient under low frequency visits. Storage space, by contrast, does not directionally shape shopkeepers' ordering behavior. While trade credits may more than offset the negative effect of low frequency visits among adopting nanostores, credit adoption remains challenging, with only 24% of nanostores assigned to the credit condition actually adopting the credit. Remarkably, not a single credit line was defaulted over the entire duration of the intervention. In terms of assortment, trade credit is mostly used to acquire nanostores' core assortment of popular, low-price items with low physical volume. We contribute to the extant literature by showing that the gesture by the supplier to extend trade credit may only partly legitimize a reduction of the visit frequency.

2 - Pay Models and Transparency on On-Demand Service Platforms

Swanand Kulkarni, Dartmouth College, Hanover, NH, United States, Basak Kalkanci, Chris Parker

On-demand service platforms have been experimenting with models to determine compensation for their workers. While some use commission- or effort-based pay models that are intuitive to workers, others, in their efforts to better match customer demand, have transitioned to models where pay is not strictly tied to effort, but depends on other, potentially exogenous factors. Platforms have also kept these pay models opaque. Workers' reactions to such practices, however, are not systematically examined or understood. Through incentivized experiments, we examine how a pay model's intuitiveness to workers, its transparency, and a change that reduces the model's intuitiveness affect workers' participation on the platform. Results show the promise and limits of pay model intuitiveness and transparency in managing worker participation.

3 - The Role of Explainable Artificial Intelligence in Influencing Loan Officer Behaviour

Stephanie Kelley, Sobey School of Business, Saint Mary's University, Halifax, NS, Canada, Anton Ovchinnikov

Little is known about whether explainable artificial intelligence (XAI) can help drive practitioners to increase the use of AI models for decision-making. We conduct a series of field and lab experiments to determine whether XAI can influence loan officer behaviour with respect to granting consumer credit.

4 - Coarsening in Human-AI Interaction

Ruru Hoong, Harvard Economics / Harvard Business School, Cambridge, MA, United States

AI signals are increasingly deployed as aids to human decision-making aids, yet various human cognitive biases can hinder their effectiveness. Coarsening at optimized thresholds—partitioning the signal space into fewer cells—can improve outcomes while keeping humans in the loop and adapting to any bias or context. Using a Bayesian persuasion framework, we determine the optimal partitioning of signals and characterize the conditions under which coarsening can help. We empirically show in a randomized experiment with loan underwriters that the provision of AI signals coarsened at the right thresholds improves overall decision-making outcomes.

MD41

Summit - 433

Randomized Experiments, A/B Tests, and Observational Studies

Invited Session

Applied Probability Society

Chair: NIAN SI, HKUST, Hong Kong, Hong Kong

Co-Chair: Zeyu Zheng, University of California, Berkeley, Berkeley, CA, United States

1 - Estimating Individual Treatment Effects Using Nonlinear Methods in Panel Data

Qingyin Ge, NUS IORA, Singapore, Singapore, Jussi Keppo, Jinglong Zhao

Synthetic control is one of the most popular methods in policy evaluation using panel data. The most classical synthetic control method proposes to use a convex combination of the untreated units to create the counterfactual for the treated units. In this paper, we propose to generalize the convex combination to nonlinear functions when doing synthetic control.

2 - Quantifying the Effect of Interference on Platform Decisions

Anushka Murthy, Stanford University, Stanford, CA, United States, Hannah Li, Ramesh Johari, Gabriel Weintraub

There has been a recent and rapidly developing literature on studying bias in A/B tests resulting from interference on marketplaces. We study the impact of bias on decision-making using a hypothesis testing viewpoint.

3 - Treatment Effect Quantiles in Stratified Randomized Experiments and Matched Observational Studies

Yongchang Su, University of Illinois, Urbana-Champaign, Urbana, IL, United States, Xinran Li

Evaluating the treatment effect has become an important topic for many applications. However, most existing literature focuses mainly on average treatment effects. When the individual effects are heavy-tailed or have outlier values, not only may the average effect not be appropriate for summarizing treatment effects, but also the conventional inference for it can be sensitive and possibly invalid due to poor large-sample approximations. In this paper we focus on quantiles of individual treatment effects, which can be more robust in the presence of extreme individual effects. Moreover, our inference for them is purely randomization-based, avoiding any distributional assumptions on the units. We first consider inference in stratified randomized experiments, extending the recent work by Caughey et al. (2021). We show that the computation of valid p-values for testing null hypotheses on quantiles of individual effects can be transformed into instances of the multiple-choice knapsack problem, which can be efficiently solved exactly or slightly conservatively. We then extend our approach to matched observational studies and propose sensitivity analysis to investigate to what extent our inference on quantiles of individual effects is robust to unmeasured confounding. The proposed randomization inference and sensitivity analysis are simultaneously valid for all quantiles of individual effects, noting that the analysis for the maximum or minimum individual effect coincides with the conventional analysis assuming constant treatment effects.

4 - Flexible sensitivity analysis for causal inference in observational studies subject to unmeasured confounding

Sizhu Lu, UC Berkeley, Berkeley, CA, United States, Peng Ding

Causal inference with observational studies often suffers from unmeasured confounding, yielding biased estimators based on the unconfoundedness assumption. Sensitivity analysis assesses how the causal conclusions change with respect to different degrees of unmeasured confounding. Most existing sensitivity analysis methods work well for specific types of statistical estimation or testing strategies.

We propose a flexible sensitivity analysis framework that can deal with commonly used inverse probability weighting, outcome regression, and doubly robust estimators simultaneously. It is based on the well-known parametrization of the selection bias as comparisons of the observed and counterfactual outcomes conditional on observed covariates. It is attractive for practical use because it only requires simple modifications of the standard estimators. Moreover, it naturally extends to many other causal inference settings, including the causal risk ratio or odds ratio, the average causal effect on the treated units, and studies with survival outcomes. We also develop an R package saci to implement our sensitivity analysis estimators.

MD42

Summit - 434

Sequential Decision-Making

Invited Session

Applied Probability Society

Chair: Yilun Chen, CUHK Shenzhen, Shenzhen, N/A

Co-Chair: Jiashuo Jiang, Hong Kong University of Science and Technology, Hong Kong, N/A

1 - Sample Complexity of Posted Pricing for a Single Item

Billy Jin, University of Chicago Booth School of Business, Chicago, IL, United States, Thomas Kesselheim, Will Ma, Sahil Singla

Selling a single item to *n* self-interested bidders is a fundamental problem in economics, where the two objectives typically considered are welfare maximization and revenue maximization. Since the optimal auctions are often impractical and do not work for sequential bidders, posted pricing auctions, where fixed prices are set for the item for different bidders, have emerged as a practical and effective alternative. This paper investigates how many samples are needed from bidders' value distributions to find near-optimal posted prices, considering both independent and correlated bidder distributions, and welfare versus revenue maximization. We obtain matching upper and lower bounds (up to logarithmic factors) on the sample complexity for all these settings.

2 - Near-Optimal Data-Driven Control for Assemble-to-Order Systems

Lun Yu, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Zhixuan Cai, Tianhu Deng, Zhaoran Wang

The literature on periodic-review assemble-to-order (ATO) systems has long assumed that the demand distribution is exactly known. In practice, this may not be true and only a limited sample of historical demand data is available. In this paper, we study data-driven ATO systems under the criteria of long-run average cost. The task is complicated by the demand learning of multiple products, synchronization between different components' ordering decisions under non-identical lead times, and allocation to different products' assembly. To the best of our knowledge, our work is the first to analyze sample average approximation (SAA) in data-driven ATO systems and the first to propose an algorithm that finds a near-optimal solution.Surprisingly, we prove that the error of SAA in optimizing over the infinite-dimensional space of policies of an ATO problem compares only logarithmically larger than that over a finite-dimensional Euclidean space.We show that the complexities associated with demand learning and high-dimensional policy optimization can be captured by properly combining a specifically designed model-free reinforcement learning (RL) algorithm with SAA, which enriches the classical inventory control literature on ATO systems. To speed up numerical computation, we prove the convexity and asymmetric Lipschitz continuity of the cost-to-go functions and design a specialized model-free RL algorithm. Using an extensive numerical study, we show that our proposed algorithm consistently and notably outperforms existing heuristic policies and an off-the-shelf RL algorithm.Our numerical results suggest that the no-holdback rule---a commonly adopted heuristic in ATO systems---may lead to a double digit performance loss.

3 - On the Limited Representational Power of Value Functions and Its Links to Statistical (in)Efficiency

David Cheikhi, Columbia University, New York, NY, United States, Daniel Russo

Identifying the trade-offs between model-based and model-free methods is a central question in reinforcement learning. Value-based methods offer substantial computational advantages and are sometimes just as statistically efficient as model-based methods. However, focusing on the core problem of policy evaluation, we show information about the transition dynamics may be impossible to represent in the space of value functions. We explore this through a series of case studies focused on structures that arises in many important problems. In several, there is no information loss and value-based methods are as statistically efficient as model based ones. In other closely-related examples, information loss is severe and value-based methods are severely outperformed. A deeper investigation points to the limitations of the representational power as the driver of the inefficiency, as opposed to failure in algorithm design.

4 - Toward Unified Lower Bounds for Statistical Estimation and Interactive Decision Making

Yunbei Xu, MIT, Cambridge, MA, United States

In this paper, we develop a unified framework for lower bound methods in statistical estimation and interactive decision making. Traditional methods such as Fano's inequality, Le Cam's method, and Assouad's lemma have been foundational for statistical estimation. However, the contemporary interactive decision-making framework necessitates new approaches, such as the Decision-Estimation Coefficient (DEC). We bridge these methodologies through a general algorithmic lower bound method, integrating the existing lower bound approaches by relaxing the separation condition using data-dependent quantile and considering an adversarial reference distribution that generates a ghost data set. Based on our general lower bound approach, we introduce a novel complexity measure in the form of a log quantile, which facilitates the derivation of new lower bounds for interactive decision making. In particular, we establish necessary and sufficient complexity measures for convex model classes, which encompass many important contextual bandit classes. Ultimately, we provide a principled unification of existing lower bound methods and enhance them in various applications.

MD43

Summit - 435

Exciting New Applications in Applied Probability

Invited Session

Applied Probability Society

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Modeling Epidemics via Queueing Theory

Jamol Pender, Cornell University, Ithaca, NY, United States

2 - Infectious Disease via Queueing Theory

Laurel Newman, Cornell ORIE, Ithaca, NY, United States

We construct and analyze a new stochastic model for infectious disease spread in service systems. Our model considers a M/M/1 queue where a proportion, p, of the arriving population is infectious and the remaining population is susceptible. When an infectious person overlaps in the queue with a susceptible customer for more than l units of time, then the susceptible person becomes infected. We provide both a personalized perspective of how an individual susceptible customer is infected and a community perspective of how one infectious customer can infect the others around them. From the personalized perspective, we derive the probability that a susceptible person becomes infected and the probability they are infected by at least k different people. From the community perspective we compute the mean and higher moments of the number of susceptible customers that are infected by one infectious individual. Finally, we use stochastic simulation to verify our analytical results and show that our new formulas accurately describe the behavior our stochastic model of infection.

3 - Understanding Bail Funds through Queueing Theory

Yidan Zhang, Cornell University, Ithaca, NY, United States, Jamol Pender

Community bail funds play a crucial role in supporting individuals unable to afford bail, significantly influencing defendants' future quality of life. We show how queueing theory is effective in modeling the dynamics of remaining money in a community bail fund and its allocation to defendants. We will demonstrate how to employ the Skorokhod Map to construct a more realistic model where defendants can be blocked from receiving funds in bail fund operations. Finally, we will discuss how our research can assist bail fund managers and stakeholders in making decisions in choosing recipients to fund.

4 - Optimal Spacing in Queues to Prevent Overlaps

Ruici Gao, Cornell University, Ithaca, NY, United States, Jamol Pender

In this paper, we investigate the probability of overlap between two customers arriving at an infinite server tandem queue with M = {2, 3, 4} stations. Understanding this probability is crucial for designing queue systems that minimize customer interaction, which is significant in a variety of realistic scenarios such as pandemic control. Our main contribution is the derivation of probability of no overlap, and the optimal spacing x between two customers required to ensure that the pair of units do not overlap within the system.

MD44

Summit - 436

Simulation Optimization and AI

Invited Session

Simulation Society

Chair: Haidong Li, University of Chinese Academy of Sciences, Beijing, N/A

Co-Chair: Gongbo Zhang, Peking University, Beijing, N/A

1 - Compass-Inspired Grid Search Algorithm: a near-Optimal Algorithm for Strongly Convex Simulation Optimization

Jianzhong Du, University of Science and Technology of China, Hefei, China, People's Republic of, Jeff Hong, Ying Zhong

In this work, a grid search algorithm is proposed to solve a class of continuous simulation optimization problems (CSO) where the objective function is strongly convex. When the solution space is one-dimensional, the algorithm exploits the nested structure of grid solutions and borrows the idea from Convergent Optimization via Most-Promising-Area Stochastic Search (COMPASS) to ensure the estimated optimal solution is locally optimal. To solve multi-dimensional CSO, the algorithm adopts the coordinate search method and changes the value of only one coordinate by solving a one-dimensional strongly convex CSO in each iteration. We find that the algorithm has an interesting link with the coordinate descent algorithm for conventional convex programming. We show that the proposed algorithm has a near-optimal convergence rate for strongly convex CSO. Numerical results show the algorithm is efficient.

2 - Optimal Sampling for Best Arm Identification with Gaussian Process Regression

Mingjie Hu, Fudan University, Shanghai, China, People's Republic of, Jie Xu, Chun-Hung Chen, Jian-Qiang Hu

We consider the fixed-budget best-arm identification (BAI) problem in the presence of spatial correlation among arms. The mean reward of each arm can only be evaluated through stochastic simulation or field experiments with heterogeneous noise. Our primary objective is to find an optimal sampling policy that considers the spatial correlation among arms to maximize the probability of correct identification. We propose using Gaussian process regression (GPR) to model the spatial correlation and develop a GPR-based framework to derive an asymptotically optimal sampling policy. Additionally, we analyze the impact of spatial correlation on the sampling policy and quantify its benefits under specific cases. We also introduce a sequential implementation algorithm and establish convergence results. Numerical experiments show that the proposed method significantly outperforms widely used benchmarks in the literature, demonstrating improved efficiency by considering spatial correlation in BAI problems.

3 - Randomized Zeroth-Order Feedback-Based Coordinate Descent Algorithm for Simulation-Embedded Problems

Ruiyang Jin, Peking University, Beijing, China, People's Republic of, Yujie Tang, Jie Song

Complex simulation processes often arise in real-world decision-making scenarios, where only an input-output oracle is available. In such situations, traditional gradient-based algorithms are ineffective as obtaining the gradient relies on access to higher-order information, which may not be observable. Zeroth-order optimization methods offer a solution by estimating gradients using the input-output oracle, which is particularly suited for high-dimensional problems. However, their sample complexity is often unsatisfactory for practical requirements. To overcome this challenge, we propose and analyze the randomized zeroth-order feedback-based coordinate descent (RZFCD) algorithm, comparing it with the conventional zeroth-order feedback-based gradient descent (ZFGD). While ZFGD applies to problems with general convex and compact feasible sets, it exhibits higher oracle complexity, typically bounded by $O(d/\epsilon^2)$. In contrast, RZFCD achieves a lower complexity bound of $O(d/\epsilon)$ but is limited to problems with box constraints. Empirical experiments are conducted to validate the performance of both algorithms.

4 - Offline Contextual Optimization with Correlated Bayesian Beliefs

Fan Zhang, ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE CHINESE ACADEMY OF SCIENCES, Beijing, China, People's Republic of

For most complex systems, the decision maker needs to make the optimal decision solution in real time with the changing exogenous information in order to minimize cost or maximize profit. While the exogenous information, commonly called contexts, will influence the performance of feasible solutions, which means that the optimal solution varies as covariates. In addition, with the inherent complexity and uncertainty pervasive in them, these complex systems are rendered intractable to analytical modeling and analysis. In this work, we investigate the offline contextual learning in simulation optimization. In offline contextual learning, the optimal design changes with the context. Therefore, the optimal design needs to be carefully determined for each possible context. Existing research has paid little attention to the impact of correlations between different designs' performances on the efficiency of offline contextual learning. In this study, we consider the Bayesian framework and utilize a multi-output Gaussian process to capture correlations between different designs' performances. We verify that considering correlations can accelerate the decay of posterior variance. Taking the KG-type strategy as an example, we construct a new Bayesian sampling policy, CIKG. Finally, we show the effectiveness of CIKG through numerical experiments.

5 - Solving Mixed Integer Linear Programs by Monte Carlo Tree Search

Gongbo Zhang, Peking University, Beijing, China, People's Republic of, Yijie Peng

Mixed Integer Linear Programs (MILPs) are powerful tools for modeling and solving combinatorial optimization problems. Solving an MILP is NP-hard due to the integrality requirement, and the branch-and-bound (B&B) algorithm is a widely used exact solution method. In this work, we explore the use of Monte Carlo Tree Search (MCTS) to guide the search within the space of branching candidate variables, aiming to efficiently find the optimal solution (if it exists) for an MILP. We adapt the Asymptotically Optimal Allocation for Trees (AOAT) algorithm, a recently proposed MCTS approach in the field of simulation and optimization, for solving MILPs. AOAT compares different nodes based on their corresponding posterior mean and posterior variance for the unknown state value. The advantages of AOAT over the commonly used Upper Confidence Bound for Trees (UCT) are that AOAT tends to explore more, achieves asymptotic optimality, and has shown better performance. Numerical results demonstrate the potential benefits of the proposed method.

MD45

Summit - 437

Invited Panel - Featured Journal - IISE Transactions on Healthcare Systems Engineering

Panel Session

Health Applications Society

Co-Chair: Hui Yang, Pennsylvania State University, University Park, PA, United States

1 - Moderator Panelist

Hui Yang, Pennsylvania State University, University Park, PA, United States

2 - Panelist

Shan Liu, University of Washington, Seattle, WA, United States

3 - Panelist

Sandra Eksioglu, University of Arkansas, Fayetteville, AR, United States

4 - Panelist

Mostafa Reisi, University of Florida, Gainesville, FL, United States

5 - Moderator Panelist

6 - Panelist

Mariel Lavieri, University of Michigan, Ann Arbor, MI, United States

MD46

Summit - 438

Data-driven Decision-making in Healthcare.

Invited Session

Health Applications Society

Chair: Vishal Ahuja, Southern Methodist University, Dallas, TX, United States

Co-Chair: Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - Personalized Treatment for Opioid use Disorder

Savannah Tang, Southern Methodist University, Dallas, TX, TX, United States, Kyra Gan, Alan Scheller-Wolf, Sridhar Tayur

Wearable devices have the potential to revolutionize treatments for *opioid use disorder* (OUD) by measuring patient responses to different treatment regimens in real-time, enabling the development of personalized treatments. A variety of wearable devices with different features, sensitivities, and costs are available. Whether such devices are practical and cost-effective to incorporate in treatments for OUD, and if so how they should be used, are critical questions. To investigate these questions, we build a finite-horizon, non-stationary *constrained partially observable Markov decision process* (CPOMDP). To facilitate the solution of our model, we provide a novel budget reformulation that finds all optimal solutions lying on the original formulation's solution convex hull. We then show that our reformulation can be solved using a binary search in conjunction with an exact POMDP algorithm. Applying those elements and using parameters estimated from past literature, we perform a numerical study to investigate the values of different wearables in OUD treatments, where we consider different levels of budget, wearable precision, and patient *treatment adherence* (TA). We find that wearables can be valuable when patients react differently to treatments across the entire population. Furthermore, this value is the largest at low or moderate budgets for patients with low or moderate TA. Outside of these settings, the marginal benefit of wearables is negligible relative to their cost.

2 - Empowering Frontline Health Workers to Tackle Stock-Outs

Amir Karimi, The University of Texas at San Antonio, San Antonio, TX, United States, Anant Mishra, Karthik Natarajan, Kingshuk Sinha, Omar Balsara, Barbara Lamphere

In low-income countries, frontline health workers are frequently tasked with the non-clinical responsibility of inventory management for which they are not adequately trained. To address these challenges, several countries have launched initiatives to empower frontline health workers in inventory management by enhancing their skills through training. Through a collaboration with JSI Research & Training Institute, Inc., we gain access to fine-grained, proprietary data from approximately 18,000 health facilities across Indonesia. Leveraging these data, we evaluate the impact of a training program for frontline health workers aimed at reducing the stock-outs of reproductive health commodities.

3 - The Impact of Virtual Visits on Ambulatory Care Utilization and Patient Health Outcomes.

Jane Iversen, The Ohio State University, Columbus, OH, United States, Yingchao Lan, Aravind Chandrasekaran

Despite the widespread introduction of virtual visits within ambulatory care, the impact that the adoption of this new care mode has on the adopting patients remains poorly understood. This study conducts a comprehensive long-term evaluation of patients' uptake of virtual visits across a wide range of medical specialties within the ambulatory care setting at a major U.S. Health System. Namely, we investigate whether virtual visit adoption affects patients' care utilization and health outcomes as reflected in their use of resource-intensive services such as emergency rooms and hospitals. Additionally, given the existing health disparities in the U.S., we seek to understand whether the effect of virtual visit adoption varies with respect to a patient's race, socioeconomic status and gender.

4 - Introducing Healthcare Delivery Equity: A Novel, Actionable Metric Providing Hospitals with Controllable Suggestions for Improvement

Jae Park, University of Central Florida, Orlando, FL, United States, Jiguang Li, Veronika Rockova, Thomas Spiegel, Daniel Adelman

While existing hospital rating systems include health equity as a metric, hospitals often find themselves unable to influence this measure due to numerous uncontrollable factors. To address this challenge from an operations management perspective, we introduce a novel metric named "healthcare delivery equity." Tailored to be under hospitals' control, it aims to facilitate the implementation of equitable care by leveraging improvements in this metric. This study assesses how equitably hospitals treat patients with the same socioeconomic status and health conditions but different races, with the goal of urging underperforming hospitals to provide equitable care. Examining the national CMS Limited Data Set data spanning 2017 to 2022, the study employs a hierarchical Bayesian approach to robustly estimate random hospital effects, considering various hospital attributes and their interactions with patient characteristics. Our Bayesian model utilizes resulting distributions unlike traditional approaches that solely rely on coefficients and classifies hospitals into two categories: equitable care hospitals and racially privileged care hospitals. This innovative rating system provides a nuanced understanding of hospitals' current practices in terms of healthcare delivery equity, thereby assisting in reshaping their decision-making processes to promote equitable care.

MD47

Summit - 439

Causal Discovery, Fairness, and Learning in Healthcare

Invited Session

Health Applications Society

Chair: Esmaeil Keyvanshokooh, Mays Business School, Texas A&M University, College Station, TX, United States

Co-Chair: Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

1 - Local Causal Discovery for Structural Evidence of Direct Discrimination

Jacqueline Maasch, Cornell Tech, New York, NY, United States, Kyra Gan, Violet Chen, Agni Orfanoudaki, Nil-Jana Akpinar, Fei Wang

Fairness is a critical objective in policy design and algorithmic decision-making. Identifying the causal pathways of unfairness requires knowledge of the underlying structural causal model, which may be incomplete or unavailable. This limits the practicality of causal fairness analysis in complex or low-knowledge domains. To mitigate this practicality gap, we advocate for the development of efficient causal discovery methods for fairness applications. To this end, we introduce local discovery for direct discrimination (LD3), a polynomial-time algorithm for recovering structural evidence of direct discrimination. Additionally, we propose a graphical criterion for identifying the weighted controlled direct effect (CDE), a qualitative measure of direct discrimination. We prove that this criterion is satisfied by the structural knowledge returned by LD3, increasing the accessibility of the weighted CDE as a causal fairness measure. We numerically validate our approach on custom synthetic data, community benchmarks, and real-world data. Taking liver transplant allocation as a case study, we highlight the potential impact of our method for modeling fairness in complex decision systems.

2 - Hybrid Top-Down Causal Discovery

Sujai Hiremath, Cornell University, Ithaca, NY, United States, Jacqueline Maasch, Mengxiao Gao, Promit Ghosal, Kyra Gan

Learning the unique directed acyclic graph (DAG) corresponding to an unknown causal model via an exact method is a challenging task. Constraint-based algorithms are limited to identifying Markov equivalence classes and often suffer from worst-case exponential runtimes. In contrast, structural equation based methods can identify a unique graph by imposing additional assumptions, but suffer from the curse of dimensionality due to high-dimensional regressions. To address these challenges, we propose a novel hybrid approach for global causal discovery in observational data. We first present a topological sorting algorithm that leverages ancestor-descendant relationships in linear structural equation models to establish a compact top-down hierarchical ordering, encoding more causal information than linear orderings produced by existing methods. We then introduce a nonparametric constraint-based algorithm that eliminates spurious edges with greater sample efficiency than current parametric methods. We provide theoretical guarantees for correctness and polynomial worst-case time complexity, supported by empirical validation on synthetic data. We demonstrate that our approach generalizes to nonlinear settings with arbitrary noise.

3 - Assessing Algorithmic Bias in Clinical Decision Support Tools for Acute Myocardial Infarction Patients: A MIMIC IV Cohort Study

Emily Garcia, Purdue University, West Lafayette, IN, United States, Zachary J. Hass

In recent years, machine learning (ML) algorithms have been increasingly integrated into clinical decision support systems (CDSS), promising to enhance decision-making in healthcare. While ML-based tools can expedite decision-making processes, there are concerns about the possibility of these tools exacerbating health disparities. Previous studies have shown that ML algorithms could inadvertently introduce bias, affecting the access and quality of care for vulnerable populations. However, there is a lack of evidence on whether ML-based CDSSs, used for early discharging of acute myocardial infarction (AMI) patients, contribute to disparities in discharge rates among vulnerable populations with AMI. This study aims to evaluate algorithmic bias and fairness in the Controlled Abciximab and Device Investigation for Lower Late Angioplasty Complications (CADILLAC) CDSS, which uses in-hospital mortality risk scores as a proxy indicator for early discharge. We examine the usage of CADILLAC on a cohort of AMI hospitalized patients to identify cases where it could be optimally used. We also propose strategies to mitigate algorithmic bias. Our findings suggest that a more patient-centered approach, considering the needs and characteristics of individual patient types, can help identify optimal patients for CDSS use. This approach could also facilitate the wider adoption of ML-based CDSSs and reduce post-discharge adverse events from impacting patients disproportionately.

4 - A Proposed Framework to Identify Upstream Sources of Bias Among Racial Subgroups in Artificial Intelligence Models

Rahul Ladhania, University of Michigan, Ann Arbor, MI, United States, Allister Ho, Karandeep Singh, Paramveer Dhillon, Chad Brummett, Anne Fernandez

Upstream biases can be understood as biases introduced into AI models prior to model training and testing. Beyond studying stratified model performance to identify these biases, we propose a framework for evaluating whether observed differences in model performance between subgroups could be due to upstream bias. Our proposed framework comprises two methodological approaches: 1) apply sub-group matching to attenuate modeling disparities and 2) explore characteristic differences between the key sub-groups to assess potential sources of model performance differences. We apply our framework to evaluate observed model performance differences between Black and White patient cohorts in two AI-based opioid use prediction models. Both models exhibited a higher AUC in Black patients as compared to White patients. We found matching Black and White patients reduced the subgroup model performance differences in our study, which was suggestive of selection bias in the unmatched cohorts. We also found that the Black patient population was more heterogeneous than the White population with respect to the distribution of predicted risk, several comorbidities, and measures of socioeconomic disadvantage. This heterogeneity is not explainable simply by our modeling procedure and likely arises from upstream factors such as historical, representation, and measurement bias, the first two of which can manifest as selection bias.

MD48

Summit - 440

Platform Operations and Retailing

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Leela Nageswaran, University of Washington, Seattle, WA, United States

Co-Chair: Soo-Haeng Cho, Carnegie Mellon University, Pittsburgh, PA, United States

1 - &Ldquo;be the Buyer&Rdquo;&Ndash;Leveraging the Wisdom of the Crowd in E-Commerce Assortment Planning

Leela Nageswaran, University of Washington, Seattle, WA, United States, Yu Kan, Uttara Ananthakrishnan

We study a new business practice of using crowdvoting, wherein a retailer first seeks input from customers on the desirability of the product and then bases the purchasing decision on their votes. We collaborated with a subscription-based apparel rental platform and obtained a proprietary dataset comprised of the platform's products and users. We find that the initiation of the service improves short- and long-term rental performance outcomes. We also investigate several potential mechanisms that drive this improvement.

2 - Two Sides of Same-Day Delivery: Evaluating the Distributional Effects of Platform Fulfillment on Third-Party Sales

Euro Bae, KAIST, Seoul, Korea, Republic of, Daegon Cho, Soo-Haeng Cho, Jin Soo Han, Yeongwoo Kim

E-commerce platforms have increasingly adopted "Fulfillment by Platform" (FBP) programs, enabling third-party sellers to delegate order fulfillment operations and improve delivery speed. Despite the prevalence of FBP programs, there is a limited understanding of how they affect sales distribution across participating FBP sellers and non-FBP sellers. This paper presents the first empirical evidence of the distributional effects of an FBP program, leveraging a quasi-experimental variation in FBP sellers' delivery speed on a fashion e-commerce platform. Specifically, the platform introduced FBP sellers' same-day delivery to customers in a specific region while keeping FBP sellers' next-day delivery in neighboring regions. Using a difference-in-differences approach, we demonstrate that the platform-driven improvement in delivery speed increases FBP sellers' sales while cannibalizing non-FBP sellers at the aggregate level. However, we find that non-FBP sellers can safeguard their sales against FBP sellers when offering vertically or horizontally differentiated products. Furthermore, our analysis of customer-level transactional data unveils the heterogeneous effects of the faster same-day delivery of the FBP program on customers: Customers who purchase from a diverse set of sellers increase their purchases from FBP sellers while reducing them from non-FBP sellers. In contrast, customers with strong loyalties to few select sellers increase their purchases from FBP sellers while maintaining purchases from non-FBP sellers. Our findings highlight the need for sellers and platforms to recognize the nuanced impacts of faster delivery on sales dynamics and the importance of product differentiation and customer loyalty building as vital countermeasures to mitigate the potential adverse effects of FBP programs.

3 - Algorithmic Pricing, Transparency, and Discrimination in the Gig Economy

Daniel Chen, Boston College, Chestnut Hill, MA, United States, Gad Allon, Kenneth Moon

Algorithms control pricing and match customers and workers in the gig economy. However, algorithms face several critiques: they lack transparency, can be biased, and can be inefficient. We empirically analyze these issues and show that variation in algorithmic pricing incentivizes workers to decline a large portion of trips in practice, significantly decreasing system efficiency. We estimate policy counterfactuals and show that minimum wage and transparent pricing policies can improve efficiency without harming fairness constraints.

4 - Capacitated Spatial Newsvendor

Jinhui Han, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Xian Yu

This work is motivated by Amazon's practice of distributing a limited supply across various locations to meet random demand. We address a spatial newsvendor problem with the aim of minimizing the total loss from transportation and inventory by optimizing supply redistribution. Employing optimal transport theory, we establish optimality conditions for different scenarios, which guide the corresponding optimal inventory reallocation strategies. Our results extend the classical newsvendor solution by incorporating spatial considerations, strategically reducing inventory levels at locations further from the supply origin. We specifically investigate both semi-discrete and discrete scenarios, offering managerial insights and developing numerical algorithms to support decision-making.

5 - Consumer Profiling via Information Design

Ruslan Momot, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Itai Fainmesser, Andrea Galeotti

Platforms use purchase histories to profile customers, create consumer segments, and disclose them to sellers. Sellers target price offers to these segments, generating new data that enables further profiling. We characterize the platform's ability to learn consumers' valuations using only information design in constructing the segments disclosed to sellers. We then evaluate the implication for market outcomes. We find that there is a threshold so that the platform cannot accurately profile consumers with valuations above it but can do so for those with valuations below it. The threshold is the seller's optimal uniform price in the no-information case. As a consequence, the use of purchase data to profile customers increases total welfare without harming consumers.

MD49

Summit - 441

Recent Advances in Retailing, Healthcare and Manufacturing Research Domains

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Anupam Agrawal, Texas A&M University, College Station, TX, United States

1 - Customized Product Manufacturing and the Role of Complexity

Anupam Agrawal, Texas A&M University, College Station, TX, United States

The production of complex, customized technology products is challenging due to the need for flexibility and frequent changes, with the impact of operational integration on productivity and cost not well understood. Our study examines a major European bus and coach manufacturer, finding that operational integration improves productivity and reduces delays, cycle time, and production costs, especially in high-complexity and high-customization scenarios. Using a novel estimation method on station-level longitudinal data, we show that real-time information sharing and inter-functional communication enhance labor and capital utilization in complex customized manufacturing settings.

2 - Going Local: Authenticity of Retailers' Buy-from-Local-Suppliers Program

Orchi Bhattacharyya, University of Minnesota, MINNEAPOLIS, MN, United States, Necati Ertekin, Mehmet Gumus, Karen Donohue

This study delves into Buy-from-Local-Suppliers (BFLS) programs, a facet of corporate social responsibility wherein grocery retailers source and showcase local food products to support local suppliers and economies. To enhance customer awareness and establish authenticity for local food products in these programs, local suppliers certify their products through third-party certifying agencies and collaborate with retailers to display local certifications on product packages and/or online product pages. Using data from a Canadian online grocery retailer that offers a BFLS program and conducting an online experiment, we assess how displaying local certifications on online product pages influences sales of local food products. Our multimethodology research provides guidance on when and how grocery retailers and local suppliers should display local certifications on online product pages.

3 - The Role of Ease-of-Return Experience on Contingent Free Shipping Policies

Nitish Jain, London Business School, London, United Kingdom, Ashish Kabra, Wedad Elmaghraby

Contingent free shipping (CFS) policies, whereby consumers get free shipping if their orders are greater than a certain amount, are commonplace. Our study documents a novel determinant of financial performance of a CFS policy: the ease-of-return experience (EOR). Our analysis indicates that our collaborating retailer would be better off selecting a (relatively) lenient CFS policy in more convenient EOR markets and a (relatively) stringent CFS policy in less convenient EOR market.

4 - Towards enabling the last-mile logistics of vaccine supply chains: learnings from covid-19 vaccine administration

Bhupinder Juneja, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Ujjal Mukherjee, Kingshuk Sinha

Vaccine administration and vaccine coverage are among the primary ways to minimize the impact of a pandemic such as COVID-19. However, policy makers face the challenge of widespread disparity in vaccine coverage across locations. While the significance of vaccine hesitancy and the associated socio-economic factors are well-acknowledged in the extant literature, the factors associated with vaccine delivery and administration – from a supply-chain efficacy perspective – are not well-understood. In this paper, we attempt to fill this gap in the literature and analyze the effect of the enablers of the last mile of vaccine supply chains with respect to three dimensions: (i) logistical connectivity, (ii) delivery infrastructure, and (iii) informational connectivity. Our results suggest that all three dimensions of the last-mile delivery and administration of vaccines improve vaccine coverage. More importantly, there exists a complementarity relationship among the logistical and informational enablers of the last-mile logistics, signifying that focusing on one of the enablers alone is not sufficient. Furthermore, we find that investing in improving the enablers of last mile of a vaccine supply chain is more beneficial for locations characterized by relatively higher prevalence of socio-economic and demographic risk factors. Taken together, the findings of the paper make a significant contribution, theoretically and practically, towards understanding the supply-chain (last-mile) enablers of delivery and administration of vaccines. In conclusion, the paper complements the literature on socio-economic factors associated with vaccine acceptance and hesitancy.

MD50

Summit - 442

Recent Advancement in Virtual Care Integration

Invited Session

MSOM: Healthcare

Chair: Yue Hu, Stanford University, Palo Alto, CA, United States

Co-Chair: Lin Zang, University of Rochester, Rochester, NY, United States

1 - The Hybrid Hospital: Balancing On-Site and Remote Hospitalization

Andrew Daw, USC Marshall Data Sciences and Operations, Los Angeles, CA, United States, Noa Zychlinski, Gal Mendelson

Hybrid hospitals offer a blend of on-site and remote hospitalization. These new healthcare models require efficient operational policies to balance costs, efficiency, and patient well-being. Our study addresses two primary operational questions: (i) how to direct patient admission and call-in policies based on individual characteristics and proximity and (ii) how to determine the optimal allocation of medical resources between these two hospitalization options. We develop a model that uses Brownian Motions to capture the patient's health evolution during hospitalization. By optimizing call-in policies, we find that remote hospitalization is cost-effective only for moderately distant patients; the call-in threshold has a non-monotonic relationship with travel time. Additionally, we find that the impact of scarce resources is reflected through simultaneous increase of both remote and on-site costs by the same value, relative to the structure under abundant resources. Lastly, we identify a non-monotonic relationship between the total medical resources and the workload allocation, depending on the recovery rates and the hospital proximity. Contrary to the widely held view that telemedicine can mitigate rural and non-rural healthcare disparities, our research suggests that on-site care may actually be more cost-effective than remote hospitalization for patients in distant locations, due to increased risks for remote patients who are called in to the hospital. This finding may be of particular concern in light of the growing number of "hospital deserts," as these communities may in fact not be well-served through remote care. Such insights can guide decision- and policy-makers in shaping future healthcare delivery and design.

2 - Recent Advancement in Virtual Care Integration

Vijaya Parameswaran, Stanford University School of Medicine, Stanford, CA, United States

Virtual care delivery, including telemedicine systems, can be conceptualized as a sociotechnical system comprising people, processes, and technologies that integrate social and technical elements to provide effective remote healthcare. This approach includes synchronous and asynchronous telehealth, remote patient monitoring, and the collection and interpretation of patient-generated health data, which have the potential to transform healthcare delivery. Since the rise of virtual care, research has highlighted positive outcomes such as high patient satisfaction, reduced travel costs, and improved patient engagement and clinical outcomes, particularly in chronic disease management. However, the utilization of such services varies significantly, influenced by factors at multiple levels within a health system, including individual preferences, clinic specialties, and organizational ideology. The effectiveness and appropriateness of virtual care also depend on patient sociodemographic characteristics, access, comfort with technology, and clinical factors such as diagnosis, clinician preferences, and clinic site. Moreover, inefficiencies arising from legacy practices, such as duplicating in-person workflows, increasing coordination burdens, and reinforcing outdated patterns, curtail the benefits of virtual care and are exacerbated by volume-focused compensation structures and organizational resistance to change. Effective virtual care integration includes identifying the 'right' patient for the 'right' modality and determining the 'right' modality for each clinic specialty, which should be based on patient demographics, technological accessibility, clinician practice patterns, clinic site, and organizational considerations. Furthermore, future research must focus on identifying specific clinical scenarios in which virtual care is most effective for maintaining access to care, guided by clinical appropriateness and utilization patterns.

3 - Sunk Cost Effect in Mitigating Service Incompletion: A Queueing Approach

Jimmy Qin, Columbia Business School, New York, NY, United States, Carri Chan, Jing Dong

Sunk-cost bias is generally recognized as a factor that can lead to suboptimal decisions. In this work, however, we investigate how sunk cost can be used as an operational lever to increase service completion rates in a congested service system. We run a controlled online experiment and find that the abandonment rate is significantly reduced for the group of participants who incur a larger sunk cost. To better capture the dynamics of service systems and their impact on customers' behavior, we study a queueing model with sunk cost and strategic customers, where customers experience a disutility of balking that is proportional to the sunk cost they incur. We characterize the equilibrium behavior of the customers, from which we further derive the optimal strategy for the service provider in terms of whether to provide real-time queue length information to customers as well as the optimal level of sunk cost to impose. Our results show that the sunk cost strategy is effective only when waiting information is provided and that using a non-zero sunk cost is optimal when the queueing system is moderately congested. Through a comprehensive numerical study, we demonstrate that implementing a non-zero sunk cost can substantially improve the throughput of the system. In addition, we reveal an asymmetric pattern in the robustness of the service provider's optimal policy which suggests that if the service provider cannot accurately estimate the customer’s sensitivity to sunk cost, using an underestimated value will give more robust performance improvements.

4 - Hospital Collaboration with Telemedicine: A Viable Solution to Address Rural Health Disparities?

Yunting shi, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Nan Liu, Sergei Savin, Guohua Wan

Telemedicine offers an innovative way of collaboration between the distinguished hospital and the ordinary hospital in a remote area. We employ a three-stage Stackelberg game to capture the dynamic interactions between the hospitals and patients, and propose a contract to incentivize telemedicine partnership, which specifies capacity allocation portion and a collaboration fee.

5 - The Impact of Information-Granularity and Prioritization on Patients' Care Modality Choice

Lin Zang, University of Rochester, Rochester, NY, United States, Yue Hu, Ricky Roet-Green, Shujing Sun

The past few years have witnessed a significant expansion in telemedicine adoption by healthcare providers. On one hand, telemedicine has the potential to increase patients' access to medical appointments. On the other hand, due to the limitations of remote diagnostic and treatment methods, telemedicine may be insufficient for patients' treatment needs and may necessitate subsequent in-person follow-up visits. To better understand this tradeoff, we model the healthcare system as a queueing network providing two types of service: telemedicine and in-person consultations. We assume that an in-person visit guarantees successful treatment, whereas a telemedicine visit may fail to meet the patient's treatment needs with a probability that is contingent on individual patient characteristics. We formulate patients' strategic choices between these care modalities as a queueing game, and characterize the game-theoretic equilibrium and the socially optimal patients' choices. We further examine how improving patients' understanding of their telemedicine suitability through predictive analytics at the online triage stage affects system performance. We find that increasing information granularity maximizes the stability region of the system but may not always be optimal in reducing the average waiting time. This limitation, however, can be overcome by simultaneously deploying a priority rule that induces the social optimum under specific conditions. Finally, leveraging real-world data from a large academic hospital in the United States, we perform a comprehensive case study that encompasses both the development of a prediction model for in-person follow-up needs and the implementation of effective information provision and prioritization strategies.

MD51

Summit - 443

Supply Chain Finance

Invited Session

MSOM: iForm

Chair: Paola Martin, Indiana University, Bloomington, IN, United States

1 - Understanding the Impact of Informational Transparency in Collaborative Ventures

Diwakar Gupta, University of Texas, Austin, TX, United States, Paola Martin, Stephen Gilbert

In many two-party collaborative ventures, such as those between a franchisor (principal) and franchisee (agent), the two firms agree to share the total realized revenue. Typically, such collaborations are subject to two informational frictions. First, while the principal advertises its contribution up front as a magnet to attract agents, it is unable to observe the effort of the agent. Second, while the two firms share the realized revenue, only the agent can observe the realized revenue. The principal may either accept the reported revenue split or choose to verify the agent’s report at a cost. A natural question that arises in this setting is whether these two informational frictions, i.e., the agent's unobserved effort and the sales revenue opacity, lower the principal’s profit. Therefore, we formulate and analyze equilibrium decisions in a game-theoretic model to tease out whether the principal would prefer full transparency if such transparency could be achieved without incurring additional costs. Contrary to intuition, we find that the principal always benefits from not observing the agent's effort. In addition, the principal may prefer positive verification costs to zero such costs provided they are not too high. There also exist parameter settings for which the agent also realizes higher profit when the principal does not observe his effort and when verification costs are positive. Thus, we show that neither of the two collaborating firms may benefit from full transparency even when it is free.

2 - Political Ideology and Global Sourcing

Jie Peng, Advanced Institute of Business, Tongji University, Shanghai, China, People's Republic of, Ben Charoenwong, Jing Wu

We study the relationship between political ideology and corporations' global sourcing strategies. Specifically, we investigate if an amplification in ideological divergence between firms and foreign governments triggers a decrease in the importation of goods from these countries. We assert that political alignment significantly impacts operational risks and regulatory compliance and consequently shapes global sourcing decisions. To empirically explore this, we utilize a difference-in-differences model and unearth a noteworthy correlation: firms that undergo an increase in ideological distance following a foreign election are found to reduce their imports from that respective country. Our findings persist across various robustness tests, thereby strengthening the reliability of our results. This research underscores the palpable effect of political ideology on international sourcing strategies, providing valuable insights for corporations maneuvering through the global political terrain.

3 - Coordination contract for time and material management in project supply chains

Xingxing Chen, University of Richmond, Richmond, VA, United States, Panos Kouvelis

We study the time and material coordination problem faced by project firms in executing project activities. Project activities often have random durations and the dependency between them makes it a challenging task to manage them. One particular challenge is that project firms need to coordinate material delivery times with suppliers in face of the activities' uncertain start time. On another front, project execution is prone to uncontrollable factors that may demand additional materials to be supplied and hence prepared for. To manage both materials' delivery time and consumption uncertainties, we develop a material coordination contract that aligns the incentives of the supplier and the project firm by accounting for both project time and cost implications. Our work bridges the gap between the standard supply chain coordination literature and the project time coordination literature.

4 - Crop Insurance Contract Design under an Intermediary

Sridhar Seshadri, University of Illinois, Champaign, IL, United States, Minyu Ye

Crop insurance is crucial for mitigating agricultural risks, while the involvement of intermediaries and government subsidies introduces complexity into the contract design. Our study revisits the problem of determining welfare-maximizing crop insurance contracts in the presence of an intermediary. We derive the optimal structure of the insurance contract, and the optimal payoffs for farmers and insurers. We also evaluate the cost of the government subsidy required to achieve the optimal welfare outcomes. By introducing the role of intermediary in the contract design, our findings offer valuable insights in designing effective insurance contracts and managing the cost of subsidies for government.

5 - Addressing Drug Shortage with Forward Contracts

Mingxuan Cui, University of Illinois Urbana-Champaign, Urbana, IL, United States, Sridhar Seshadri, Qiong Wang

Drug shortage poses significant challenge for today's healthcare systems. Inadequate capacity, raw material shortage, and production disruption are major contributing factors to the problem. Addressing these issues requires business innovation. To this end, we consider an emerging drug supply chain model in which a third-party intermediary offers producers and buyers forward contracts to align incentives and reduce uncertainty. The business process is formulated as a multi-stage game that involves manufacturers' capacity decisions and subsequent competitions a la Cournot. As a significant feature of the model, random yield of capacity investment not only raises profit risks but also incentivizes manufacturers to take opportunistic behaviors to benefit from drug shortage. By comparing equilibrium outcomes of the game with and without the intermediary, we demonstrate the important role that third-party forward contracts can play to reduce drug shortage.

MD52

Summit - 444

Pricing and Learning in Retail Operations

Invited Session

MSOM: Service Operations

Chair: Kostas Bimpikis, Stanford University, Stanford, CA, United States

Co-Chair: Vasiliki Kostami, HEC Paris, Jouy-En-Josas, France

1 - Routing in Vendor-Managed Inventory Supply Chains: Efficiency and Assortment Quality

Kostas Bimpikis, Stanford University, Stanford, CA, United States, Salomon Wollenstein-Betech

While the theoretical literature on the optimal and fair allocation of limited inventory to a set of retailers is extensive, there has been little empirical research to explore how inventory allocation strategies perform in real-world supply chains and their impact to downstream retailers. Using a proprietary panel dataset from a major CPG company in Mexico, we uncover a relationship between the order retailers were allocated inventory and the service quality they receive. Specifically, our findings illustrate significant disparities on both the inventory level and the product assortment between retailers depending on their position in the delivery route from the distribution center, which determines the order by which they were allocated inventory. Retailers that receive inventory last tend to receive more inventory of the best-selling products and typically carry higher priced SKUs. This finding had a direct impact on the retailers' sales, profits, and returns.

2 - Managing Acquisition and Retention Dynamics under Demand Learning and Capacity Constraints

Seyed Seyed Ghafouri, London Business School, London, United Kingdom, Oded Koenigsberg, Vasiliki Kostami

This research addresses the critical challenge of managing customer acquisition and retention in environments characterized by demand uncertainty and capacity constraints. We develop a dynamic model that incorporates stochastic approximation techniques to optimize pricing as well as acquisition and retention strategies by leveraging real-time interaction data and capacity constraints in decision-making processes. We demonstrate that our algorithm achieves optimal performance with theoretical guarantees, matching the best achievable regret in the online learning context. Our findings offer valuable insights for marketing and operations managers, addressing the joint interface in customer relationship management.

3 - Incentives in Online Gaming: Optimal Policy Design with Dynamic Causal Machine Learning

Zhikun Lu, Emory University, Decatur, GA, United States, Ruomeng Cui

We study the trade-off between AI and human decisions on reward design under static and dynamic settings in online gaming. We first document the empirical evidence and then develop an optimization framework for optimally assigning treatment under a dynamic setting to maximize user engagement.

4 - Market Fragmentation and Inefficiencies in Maritime Shipping

Giacomo Mantegazza, Amazon, Seattle, WA, United States, Kostas Bimpikis, Salomon Wollenstein-Betech

Maritime transportation is essential for global supply chains, but ballasting—vessels traveling without cargo—imposes significant economic and environmental costs. This paper focuses on the oil transportation industry, where about half of the total traveled miles are sailed empty, and illustrates that market fragmentation is the most important cause to ballasting after demand and supply imbalances, accounting for 17-20% of the total. We find that it is possible to reduce carbon emissions associated with ballasting by as much as 15% by consolidating the market into small shipping pools, i.e., sets of vessels controlled by a single operator, without raising concerns about excessive market power. Consolidation improves utilization due to larger pools coordinating their vessels more efficiently and expanding the set of ports they serve, which reduces the need for vessel relocations. At a broader level, this work shows the extent of the sustainability gains that can be obtained solely by organizing efficiently the resources available in today's supply chains.

MD53

Summit - 445

Innovating the Supply Chain: Scientific Advancements at Chewy

Invited Session

MSOM: Supply Chain

Chair: Eric Huang, Chewy, Bellevue, WA, United States

1 - Predicting Purchase Order Fulfillment: Integrating Machine Learning with Supply Chain Optimization

Eric Huang, Chewy, Bellevue, WA, United States, Chong Sun

To optimize the creation of Procurement Orders (POs), we should have an accurate measurement of how many units from the on-order POs will arrive. This motivates us to develop a model to better predict arrival quantity. Further investigation shows that approximately 97% of items in on-order POs arrive either in full (order quantity equals received quantity) or not at all and 3% arrive with a partial quantity. We thus decided to first develop a binary classification machine learning (ML) model that predicts if an item in an on-order PO will arrive. Compared to the multiclass classification or regression models, the binary classification model is easier to develop and test. Should the portion of partial received quantity arise, we will likely develop a different model that predicts the filled quantity directly. We call the current binary classification model, filled-quantity model (FQM), and it is based on a XGBoost method. Using these predictions, we can update the on-hand inventory estimate for an item during its expected receiving week. Once integrated into our PO creation system, we can make more informed decisions on new orders. We also conducted an entitlement study of the FQM based on a simulation method we derived without expensive integration and backtests. This helps us to bring the leadership onboard in implementing our model.

2 - Reinforcement learning based buying decisions in pharmacy business

Baris Kacar, Chewy, Bellevue, WA, United States, Chong Sun

The pharmacy business is complex, as there are certain rules and regulations that impact how and where products can be stored and shipped from. Fulfilling pharmacy orders in a short time is the highest priority and important for our customers, so it is essential that our fulfillment carries the right amount of products in the right places. One level above is the decision to order the correct number of products from our vendors, which keeps the pharmacy locations replenished regularly and enables them to serve with approximately 100% service levels. With the highly complex structure of decision-making in the face of stochastic lead times, fill rate reliability, and other supply chain constraints such as capacity, we explore reinforcement learning in buying decisions for the pharmacy business at Chewy. We compare different strategies in reinforcement learning and show numerical results to highlight how it can be used as an alternative approach to other stochastic or MILP-based models.

3 - Scaling Order Routing Optimization: Integrating MILP with GNN

Venkatesh Manohar, Chewy, Plantation, FL, United States, ChanSeng Pun, Gautham Sunder

Mixed Integer Linear Programming (MILP) is widely adopted for numerous discrete optimization tasks. However, lack of scalability limits their application in large scale use cases such as e-commerce delivery optimization models, which necessitates exploration of alternate methods to solve them efficiently. This study addresses the order routing problem modeled using MILP with an objective that minimizes the overall fulfillment cost and delivery time. We pre-solve the problem using a Graph Neural Network (GNN) to improve the scalability of the problem. The remaining problem is solved following the traditional approach. Our preliminary results tested using Chewy supply chain network suggest that the hybrid approach, which integrates the use of GNN and MILP, addresses the scalability limitation with minimal optimality tradeoffs

MD54

Summit - 446

Operations for the Future of Work

Panel Session

MSOM:Technology, Innovation, and Entrepreneurship

Co-Chair: Park Sinchaisri, University of California Berkeley, Berkeley, CA, 94612, United States

1 - From Factories to Streets? Investigating the Impact of Ridesharing Services on Labor Turnover in U.S. Manufacturing

Yasaman Asayesh, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Anant Mishra

How does Uber's entry into ride-sharing unexpectedly drive labor turnover in small manufacturing firms and what it means for the future of work?

2 - Uber Your Cooking: The Sharing Economy Operations of a Ghost Kitchen Platform

Feihong Hu, University of Texas at Austin, Austin, TX, United States, Junyu Cao, Wei Qi

How are ghost kitchens transforming urban cooking, optimizing profitability through chef specialization while overcoming challenges of wait times and delivery costs?

3 - Managing Competitive Multi-agent Environments with Recommendation Algorithms

Bryce Mclaughlin, University of Pennsylvania, Philadelphia, PA, United States, Park Sinchaisri

Managing Competitive Multi-agent Environments with Recommendation Algorithms" - How do workers strategically comply with recommendations and how does competition shape algorithm use?

4 - Algorithm Reliance, Fast and Slow

Clare Snyder, Michigan Ross, Ann Arbor, MI, United States

How do algorithm quality and system load impact both decision-making speed and service quality?

5 - Last-mile Delivery in Health Care: Drone Delivery for Blood Products in Rwanda

Harriet Jeon, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

How does drone adoption for blood product deliveries in Rwandan hospitals impact inventory management and health outcomes?

MD55

Summit - 447

Doing good with OR: Responding to the needs of the vulnerable population

Invited Session

Public Sector OR

Chair: Faith Idoko, North Carolina Agricultural and Technical State University, Greensboro, NC, United States

1 - Supply Chain Resilience in the Face of Uncertainty: A Food Bank Case Study

Faith Idoko, North Carolina Agricultural and Technical State University, Greensboro, NC, United States, Chrysafis Vogiatzis, Lauren Davis

Humanitarian logistics has played an essential role in disaster management over the years. However, anticipating and preparing for the aftermath of these events can be very challenging due to its unpredictability. When such situations arise, there is often a notable increase in the number of food insecure individuals within the impacted area. To curb potential food crises after the disaster, food banks need to better prepare by evaluating network vulnerabilities beforehand. Recognizing the magnitude of the disaster, social vulnerability and resilience indices of the affected area can guide distribution efforts in meeting demand effectively and on time. Additionally, possible changes to this anticipated demand, as well as the available supply and physical capacity of the food bank must also be considered. These variabilities can stem from residents' evacuation behavior, or the trajectory and magnitude of the disaster. We propose a decision support approach which accounts for potential uncertainties that may impact preparations made by the food bank, or their ability to respond to the needs of the food insecure during dire times. The proposed approach in this study boosts food bank disaster preparedness, ensuring the food bank still accomplishes its main goal of serving the food insecure population despite the limitations imposed on its network by the disaster. Findings from this work can be adapted to fit food banks of various sizes with comparable network structures while also accounting for varying disaster types.

2 - Collaborative Prepositioning with Postponement and Stock Sharing

Lamia Gulnur Kasap, Ozyegin University, Istanbul, Turkey, Birce Adsanver, Burcu Balcik

We evaluate the impacts of postponement and stock-sharing strategies among humanitarian agencies operating within regional and country-level depots for prepositioning supplies. Our proposed system involves agencies stocking supplies in both branded (final product) and unbranded (shareable standard) forms at the regional depot, while limiting stocks to branded form only at country depots for faster response. To minimize expected response time under fixed total base stock constraints in this setting, we develop a two-stage stochastic model that determines the optimal inventory levels of branded and unbranded items for each agency across these depots. Our model incorporates constraints for facilitating practical stock sharing among agencies and stock mobilization from each agency to countries within its response region. We test our approach using historical hurricane scenarios from the Caribbean region and data from agencies. We demonstrate substantial savings in fill rate, response time, and inventory utilization.

3 - Co-designing solutions to societal issues with lived experts

Kayse Lee Maass, Northeastern University, Boston, MA, United States, Baris Tezcan, Geri Dimas, Yaren Bilge Kaya, Thomas Sharkey, Renata Konrad, Andrew Trapp, Lauren Martin, Kelle Barrick

Operations Research has the potential to help address many of the pressing societal issues facing our world today that increase the vulnerability of people to experience harm. But, how do we know what the needs of people in vulnerable positions actually are? This presentation describes the ways we have partnered with people with lived experience (e.g., homeless youth, human trafficking survivors) to understand their needs, rather than presuming knowledge of their needs without their involvement. The goal of building these community-based research partnerships is to move away from extractive research *on* vulnerable communities to co-creating research and solutions in partnership *with* people with lived experience. Such an approach also helps to understand and mitigate any unintended negative consequences that may come from implementing the policies, technologies, or operational changes suggested by the models. The audience will leave with an understanding of why this approach is necessary and actionable tips for how to create and sustain these partnerships.

4 - Assessing the Relationship between Social Vulnerability and Assessed Housing Damage in Hurricane Michael

Christopher Zobel, Virginia Tech, Blacksburg, VA, United States, Andrew Arnette, Louis Luangkesorn, Michael Whitehead

We discuss a project that combines three large-scale different data sets to better understand the interplay between social vulnerability and housing damage that occurred in Florida as a result of Hurricane Michael. The damage data is drawn from an American Red Cross post-disaster damage assessment. Centers for Disease Control and Prevention / Agency for Toxic Substances and Disease Registry (CDC/ATSDR) social vulnerability data is then included at the county level, where it is subdivided into its four component themes. Detailed weather data is also incorporated into the analysis at the county level in order to examine the relative strength of the storm in each affected area. We discuss preliminary results that confirm the hypothesis that different aspects of social vulnerability, as captured in the CDC/ATSDR themes, have differential impacts on the overall housing damages that resulted from the storm. This serves to further emphasize the importance of addressing the needs of vulnerable populations in order to improve their resilience to natural disasters.

5 - Supply chain network design for not-for-profit food distribution

Christine Nguyen, Northern Illinois University, Dekalb, IL, United States, Gabriela Ibarra, Michael Hewitt

This study describes a two-stage stochastic problem that can be formulated and solved to design a supply chain for a third party logistics (3PL) provider that equitably supports the goals and missions of food banks and food pantries. The primary purpose of this model is to help a 3PL determine which consolidation centers to open to equitably balance food distribution across food banks, and subsequently, food pantries. A secondary decision is to determine the size of the fleet of vehicles owned by the 3PL to hedge against uncertainty in the transportation spot market. In this study, an algorithm is developed and compared with CPLEX results.

MD56

Summit - 448

Life as a Ph.D. Student

Panel Session

Minority Issues Forum

Co-Chair: Jiahui "Gary" Luo, Dartmouth College, Hanover, NH, United States

1 - Moderator

Jiahui "Gary" Luo, Dartmouth College, Hanover, NH, United States

2 - Panelist

Kara Combs, Air Force Research Laboratory, Wright-Patterson AFB, OH, United States

3 - Panelist

Edikan Udofia, Colorado School of MInes, Golden, CO, United States

4 - Panelist

Amaya McNealey, Georgia Institute of Technology, Madison, AL, United States

5 - Panelist

Zijin Zhang, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

MD57

Summit - Terrace Suite 1

Healthcare Analytics and Operations

Invited Session

Health Applications Society

Chair: Jin Qi, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

Co-Chair: Aiqi Zhang, Wilfrid Laurier University, Waterloo, ON, Canada

1 - Dynamic Capacity Management for Deferred Surgeries: a Case Study on Hernia

Eojin Han, University of Notre Dame, Notre Dame, IN, United States, Kartikey Sharma, Kristian Singh, Omid Nohadani

We studied 4 years of medical insurance claims of more than 15,000 hernia patients in the United States and observed the presence of uncertainty in surgery demand and patient departure. The endogeneity of these uncertainties to hospital operations renders existing capacity management approaches inapplicable. To this end, we develop two approaches based on robust and distributionally robust optimization. These methods offer sizable improvements over alternative methods for hernia patients. We also discuss multiple operational insights obtained from extensive sensitivity analysis.

2 - Distributionally Robust Group Testing with Correlation Information

Jin Qi, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Zhuoyu Long, Yu SUN, Aiqi Zhang

Motivated by the need for more efficient and reliable methods of group testing during widespread infectious outbreaks, such as the COVID-19 pandemic, this paper introduces a novel operational improvement to Dorfman's widely-used group testing procedure. Our method minimizes a weighted sum of tests and misclassifications, predicated on the known prevalence rates and interindividual Pearson correlation coefficients. Recognizing the inherent ambiguity in the population-level probability of infections that arises from correlations, our approach leverages a distributionally robust optimization (DRO) framework to counteract the worst-case probability distribution. In fully-correlated cases, where each pair of subjects are equally correlated, we establish uniform group sizes and connect our analysis to Nash equilibrium principles. Larger testing groups are generally favored under high correlation, whereas individual testing becomes optimal under high prevalence. In partially-correlated cases, where the population is formed by several intra-correlated but inter-independent clusters, we highlight the effectiveness of mixed-cluster testing strategies, particularly at lower levels of prevalence and correlation. Conversely, scenarios with high prevalence or high correlation tend to favor individual testing or same-cluster pooling. For both fully- and partially-correlated cases, we develop polynomial-time solutions and conduct a thorough exploration on the change of optimal pooling strategy as a function of imperfect tests. We demonstrate the benefits of adopting the DRO framework through a comprehensive comparison with stochastic alternatives, and we illustrate the significant impact of considering correlated infections through a case study on a COVID-19 dataset from Hong Kong.

3 - Satisficing Measure for Extubation Decision Analytics

Jingui Xie, Technical University of Munich, Heilbronn, Germany, Zhiyuan Lou

The process of deciding when to extubate patients who rely on mechanical ventilation for breathing is critical in intensive care settings. This research focuses on determining the optimal timing for extubation by taking into account the risks related to the variability of outcomes. This study introduces a satisficing Markov decision process (MDP) that employs a satisficing measure to blend risk tolerance with a decision-maker-specified target. It unveils the threshold-based nature of the optimal policy, the submodularity of the value function, and a policy structure that is aware of risk. We identify three scenarios of risk-sensitive strategies for decision-making in comparison to the nominal MDP: an aggressive policy when terminal costs are deterministic, a conservative policy, and a mixed strategy that transitions from conservative to aggressive actions. Numerical experiments demonstrate our model's effectiveness, revealing a balance between expected costs and risk profiles (e.g., approximately 1.15% increase in expected cost in exchange for a 22.13% reduction in conditional expected shortage). These findings highlight a key balance between the risk of immediate termination and the ongoing risk associated with continuing treatment. Utilizing our model enables managers to significantly diminish the risk of shortages with a minimal increase in expected costs.

4 - Queueing Causal Models: Comparative Analytics in Queueing Systems

Zhenghang Xu, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Opher Baron, Dmitry Krass, Arik Senderovich, Mark Van der Laan

Simulation is a powerful tool for comparative analysis of queueing models. With expert knowledge of underlying system structure, simulator can be constructed to predict intervention effects. However, such manual construction is time- and skill-demanding. It could also be subjective - if expert failed to note an important system feature (e.g. different customer types receiving different service priorities), the model will not be accurate. As an alternative, we propose a data-driven representation of system building blocks, justified by G-computation formula. We describe the queueing data generation process with causal graphs and apply machine learning models to fit the structural equations. Through numerical experiments, we show that this approach can replace the explicit queueing dynamics and capture intervention effect in overtake-free queues.

MD58

Summit - Terrace Suite 2

OR Applications to Improve the Quality of Life of Individuals with Disabilities

Invited Session

Health Applications Society

Chair: Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Developing Score-Based Risk Assessment for Autism Using Supersparse Linear Integer Model

Yu-Hsin Chen, Penn State University, University Park, PA, United States, Whitney Guthrie, Qiushi Chen

Early diagnosis is crucial to the treatment outcome of Autism Spectrum Disorder (ASD), a developmental disorder that affects 1 in 36 children in the US. The Modified Checklist for Autism in Toddlers (M-CHAT) is the most commonly used screening tool for ASD-specific screening during pediatric well-child visits. However, its effectiveness is limited due to concerns about its accuracy. Although previous research has demonstrated that machine learning models using electronic health record (EHR) data can enhance ASD risk predictions, these models often involve complex computations and are challenging to integrate into clinical practice because they do not incorporate M-CHAT results directly and are difficult to interpret. To address these issues, we have applied a Supersparse Linear Integer Model (SLIM) to develop a score-based risk assessment system. This new model complements the existing M-CHAT by simplifying the calculation process and enhancing the overall accuracy of ASD screening. This approach promises to make advanced risk prediction models more accessible and practical for routine clinical use.

2 - Data Analytics Framework to Evaluate the Risk of Long-Term Side Effects of Psychotropic Polypharmacy in Autism Spectrum Disorder

Paula Andrea Jaimes Buitron, The University of Texas at San Antonio, San Antonio, TX, United States, Leslie Neely, Melissa Svoboda, Carolina Vivas-Valencia

Children with autism spectrum disorder (ASD), particularly those severely affected, are at high risk for challenging behaviors like physical aggression, self-injury, and property destruction. Without proper intervention, these behaviors can worsen over time. It is estimated that up to

65% of individuals with challenging behavior are treated with psychotropic medications, and 35% of them are exposed to polypharmacy, defined as the utilization of two or more prescriptions concurrently for at least 30 days. Psychotropic medications are drugs used to treat various psychiatric and neurological disorders that affect brain function, behavior, mood, and thoughts. Even though the use of psychotropic medications has become more common in the ASD population, there is still a significant gap in our understanding of their utilization in clinical practice. This study aims to determine the utilization and risk of long-term side effects, including metabolic syndromes associated with psychotropic polypharmacy among individuals with ASD. Leveraging a national cohort of commercially insured patients with ASD, we first investigated changes in the utilization of psychotropic medications intended to manage challenging behavior among children, adolescents, and adults with ASD in the US. Second, we conducted a survival analysis to determine the probability of individuals with ASD who did not face metabolic syndrome during the treatment exposure to psychotropic interventions from 2012 to 2021. This study offers insight into the use of psychotropic polypharmacy and the associated risks among various ASD demographic groups, including age, sex, and place of residence, contributing to the development of evidence-based care for individuals with ASD.

3 - Optimizing Patient Outcomes Through Generative Graph Neural Networks for Predictive Modeling of Multi-Chronic Conditions

Julian Carvajal Rico, The University of Texas at San Antonio, San Antonio, TX, United States, Adel Alaeddini, Syed Hasib Akhter Faruqui, Susan P Fisher-Hoch, Joseph B Fisher-Hoch

Predicting the emergence of multiple chronic conditions (MCC) is crucial for early intervention and personalized healthcare, as MCC significantly impacts patient outcomes and healthcare costs. Graph neural networks (GNNs) are effective methods for modeling complex graph data, such as those found in MCC. However, a significant challenge with GNNs is their reliance on an existing graph structure, which is not readily available for MCC.

To address this challenge, we propose a novel generative framework for GNNs that constructs a representative underlying graph structure by utilizing the distribution of the data to enhance predictive analytics for MCC. Our framework employs a graph variational autoencoder (GVAE) to capture the complex relationships in patient data. This allows for a comprehensive understanding of individual health trajectories and facilitates the creation of diverse patient stochastic similarity graphs while preserving the original feature set. These variations of patient stochastic similarity graphs, generated from the GVAE decoder, are then processed by a GNN.

Next, a contextual Bandit is designed to evaluate the stochastically generated graphs and identify the best-performing graph for the GNN model iteratively until model convergence. We validate the performance of the proposed contextual Bandit algorithm against e-Greedy and multi-armed Bandit algorithms on a large cohort (n = 1,592) of patients with MCC. These advancements highlight the potential of the proposed approach to transform predictive healthcare analytics, enabling a more personalized and proactive approach to MCC management.

4 - A Ride-Sharing Platform for Improved Healthcare Access by People with Disabilities

Omar Abbaas, The University of Texas at San Antonio, San Antonio, TX, United States

This study addresses the transportation barrier faced by People with Disabilities (PWD) accessing healthcare services. We propose a novel approach through the development of a ride-sharing program facilitated by a mathematical model and a matching algorithm. This program will connect PWD with volunteer drivers and other community members who can provide transportation assistance. The matching process will consider various factors including the type of disability, geographic location, preferred language, and service type (paid or volunteer). By incorporating these details, the program aims to create efficient and comfortable transportation solutions that cater to the specific needs of PWD. To ensure effectiveness, the program's functionality will be validated using simulation tools. This will allow us to generate alternative recommendations, further optimizing the matching process and ultimately improving healthcare access for PWD.

MD59

Summit - Ballroom 1

2024 MSOM Distinguished Fellow Talks

Award Session

Manufacturing and Service Operations Management (MSOM)

Chair: Lingxiu Dong, Washington University in St. Louis, Saint Louis, MO, United States

1 - The Story of My Research

Mor Armony, New York University, New York, NY, United States

?

2 - Operations in the Age of the Sharing Economy

Saif Benjaafar, University of Michigan, Ann Arbor, MI, United States

?

3 - Empirical Research in OM: A Recipe for Avoiding Pitfalls and Creating Lasting Impact

Vishal Gaur, Cornell University, Ithaca, NY, United States

?

MD60

Summit - Ballroom 2

AI at Amazon Last-Mile

Invited Session

The Practice Section of INFORMS

Co-Chair: Abhilasha Katariya, Amazon, Issaquah, United States

1 - Revolutionizing Last Mile Delivery through Channel Topology Optimization

Gokce Kahvecioglu, Amazon, Bothel, WA, United States, Jin Ye

The expansion of Amazon Last Mile Delivery is one of the key enablers for the continued growth of Amazon's consumer business. Sustaining this growth is dependent on our ability to move beyond our traditional delivery channels, Delivery Service Provider (DSP) and Amazon Flex (Flex), to leverage alternative channels such as Hub Delivery Partners (Hub), walkers and cargo electric bikes (E-Bikes).These alternative delivery methods can help Amazon to overcome capacity constraints, real estate limitations, labor shortages, improve the delivery cost and speed and also support Amazon's sustainability goals. AMZL needs to determine the most suitable combination of delivery methods for a given geography to ensure optimal efficiency and lower cost, yet to meet customer demand and delivery speed, and make progress towards its sustainability commitments. This paper aims to develop an optimized multi-modal transportation model - Channel Topology Optimization (CTO) and proposes an expansion strategy at different phases of planning process of last-mile delivery.

2 - Large Scale Events

Selin Tosun, Amazon, Seattle, WA, United States, Jin Ye, Leah Riley

Large-scale events (LSEs) are not only impacting our scheduling and delivery quality but also many other aspects of pool health. In 2021 in the US, there were deliveries affected by these events, and led to drops in the pool health. Exempting these defects is currently handled manually, and some of these events are not processed on time. Anomaly-based approaches are not able to catch LSE events as effectively due to the non-linear nature of the problem. Here, we present a novel approach on automating LSEs detections by the systemic observations in the environmental factors and defect densities by XGBoost model for classifying LSEs. We are able to auto-capture more than 83% of the events accurately and immediately after the block's defects are posted, resulting in near-real-time Anomaly detection. The initial observations in real operations showed improvement on the timeliness metric to 95%, by automation and dramatically reduces the number of manual processes; the full extent is still being calculated. The LSE-Anomaly detection does not only improve the timeliness of the exemptions, but also standardizes the cases by reducing the inconsistencies coming from the manual investigations and reduces the dependencies on the event exemption intake requests. We see additional benefits of this modeling approach in wider set of applications in the Evaluations space.

3 - Optimizing Last Mile Delivery with Machine Learning: Amazon's Infrastructure and Scaling Strategies

Xiaodong Lan, Amazon, Inc, Seattle, WA, United States, Eliot Hijano, Ram Thiruveedhi, Abhilasha Katariya, Rohit Malshe

This research presents a novel approach leveraging Amazon Web Services (AWS) to enhance forecasting for Amazon's last-mile delivery operations. We describe our methodology for data collection, feature engineering, and developing cutting-edge machine learning models powered by AWS services. By integrating these models into decision-making, we demonstrate the potential for significant cost saving. We provide insights into our debugging and parameter tuning strategies facilitated by the AWS ecosystem, highlighting the value of cloud infrastructure in building high-performance forecasting models. The findings offer practical implications for logistics organizations seeking to optimize operations through data-driven decision-making supported by advanced analytics and cloud computing.

4 - Cold-start Forecasting for Special Events

Arkajyoti Misra, Amazon.com, Bellevue, WA, United States, Abhilasha Katariya, Ram Thiruveedhi

Accurate forecasting for rare events like high-demand, or special sales periods is hampered by limited data availability, leading to suboptimal planning and increased costs. This paper explores multiple solutions to address the cold-start problem, including imputing from relevant scenarios with appropriate adjustments and appending historical data to augment sparse datasets. We evaluate the effectiveness of these approaches in improving forecasting accuracy for key operational metrics during such special events, within the Amazon context. The proposed methodologies enable better capacity planning, mitigating risks associated with understaffing or overstaffing, and ultimately driving cost optimization.

MD61

Summit - Ballroom 3

AI and Generative AI in Analytics Practice

Panel Session

The Practice Section of INFORMS

Co-Chair: Segev Wasserkrug, IBM Research and Faculty of Data and Decision Sciences Technion, Technion, Israel

1 - Panelist

Jonathan Helm, Indiana University, Bloomington, IN, United States

2 - Panelist

Nobel Niraula, Boeing, Huntsville, AL, United States

3 - Panelist

Louis Luangkesorn, Highmark Health, Pittsburgh, PA, United States

4 - Panelist

Leonard Boussioux, UW Foster, Harvard, Seattle, WA, United States

5 - Panelist

Patricia Neri, UL Solutions, Cary, NC, United States

MD62

Summit - Signature Room

Randomized Rounding Approaches to Online Allocation, Sequencing, and Matching

Invited Session

TutORial

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Randomized Rounding Approaches to Online Allocation, Sequencing, and Matching

Will Ma, Columbia University, New York, NY, United States

Randomized rounding is a technique that was originally used to approximate hard offline discrete optimization problems from a mathematical programming relaxation. Since then it has also been used to approximately solve sequential stochastic optimization problems, overcoming the curse of dimensionality. To elaborate, one first writes a tractable linear programming relaxation that prescribes probabilities with which actions should be taken. Rounding then designs a (randomized) online policy that approximately preserves all of these probabilities, with the challenge being that the online policy faces hard constraints, whereas the prescribed probabilities only have to satisfy these constraints in expectation. Moreover, unlike classical randomized rounding for offline problems, the online policy's actions unfold sequentially over time, interspersed by uncontrollable stochastic realizations that affect the feasibility of future actions. This tutorial provides an introduction for using randomized rounding to design online policies,

MD63

Regency - 601

ICS Award Session

Award Session

Computing Society

Chair: Thiago Serra, Bucknell University, Lewisburg, PA, United States

Co-Chair: Merve Bodur, University of Edinburgh, Edinburgh, United Kingdom

1 - ICS Harvey J. Greenberg Research Award: Self-adapting Network Relaxations for Weakly Coupled Markov Decision Processes

Andre Cire, University of Toronto, Toronto, ON, Canada, Selva Nadarajah

To be entered.

2 - ICS Student Paper Award: Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods

Shuvomoy Das Gupta, Columbia IEOR, NYC, NY, United States

Coming soon

3 - ICS Prize: Riemannian Optimization

Shiqian Ma, Rice University, Houston, TX, United States, Krishnakumar Balasubramanian, Shixiang Chen, Jiaxiang Li, Anthony So, Tong Zhang

Coming soon

MD64

Regency - 602

Analytics in Digital Platforms

Invited Session

Social Media Analytics

Chair: Subodha Kumar, Fox School of Business, Temple University, Philadelphia, PA, United States

Co-Chair: Rakesh Mallipeddi, Th Ohio State University, Columbus, OH, United States

1 - Incorporating Genetic Data into Outcome Prediction and Treatment Prescription for Ovarian Cancer

Neelkanth Bardhan, Massachusetts Institute of Technology, Cambridge, MA, United States, Alkiviadis Mertzios, Matea Gjika, Xidan Xu, Samayita Guha, Subodha Kumar, Angela Belcher, Georgia Perakis

Ovarian Cancer (OC) is the leading cause of gynecological cancer-related deaths among women, with a 5-year survival rate of about 50% across all stages. To improve patient prognosis, there is an unmet clinical need for advanced Machine Learning (ML) models that utilize genetic data to improve OC outcome predictions and treatment recommendations. Our study addresses this shortcoming by developing new techniques to incorporate genetic information into treatment planning for OC, using data from the Prostate, Lung, Colorectal, and Ovarian Cancer (PLCO) trial. In this work, we transform genotype sequencing data into a tabular format suitable for ML models. By combining this genetic data with clinical parameters and recorded treatment regimens, we estimate treatment effects on mortality and recommend optimal treatments. Compared to models using clinical parameters only, we achieve improved efficacy of our predictive model by including genetic data, as demonstrated by an increased accuracy of prediction. Our ML models enable oncologists to make more informed decisions, potentially leading to improved prognosis and better survival outcomes for OC patients.

2 - On the Vertical Spillover Effect of Online Ratings on Platform Competition

Yulia Vorotyntseva, University of St. Thomas, Minneapolis, MN, United States, Aleksi Aaltonen, Subodha Kumar, Paul Pavlou

The standard 5-star rating system is a key source of information for consumers trying to decide where to eat, what products to buy, or which doctor to visit. Yet, a vendor may have different average rating on different platforms and, consequently, a consumer may observe the same vendor with different average ratings across platforms. Higher average ratings are known to increase vendor sales, but we do not know how differences in average ratings across platforms may affect platform competition.We study the impact of the difference in average ratings for the same vendor across platforms on the choice of a platform in a setting in which vendors offer their products (multihome) on several platforms and consumers can easily switch between platforms. Using incentive-aligned and stated-preferences experiments in a food delivery setting, we show that, all other things being equal, consumers prefer to purchase from a platform that displays a higher average rating for the same vendor---even if the difference in ratings would not signal a real difference in the quality of the product or purchasing experience from the same vendor between the platforms. We explain this by a vertical spillover effect by which consumers allow vendor ratings to bias their platform choice. The results have implications to how platform owners and policymakers should govern the provision of online ratings. When both vendors and consumers multihome across platforms, the management of each platform must govern its rating system strategically to avoid offering a virtual showroom to competitors.

3 - Are Smart Contracts Smart Enough? A Game Theoretic Approach for Implementing Blockchain-Enabled Group Purchasing

Amirhossein Ranjbar, Fox School of Business, Temple University, Philadelphia, PA, United States, Subodha Kumar

While improved supply chain transparency stemming from blockchain adoption can help supply chain members enhance their operational decisions, sharing information can come at the expense of competitive disadvantage. Consequently, firms often face challenges when considering blockchain technology adoption. In this study, we explore the optimal decisions of a single supplier and two competing retailers who face uncertain demand at the end market. Retailers do not know each other's demand, resulting in bilateral asymmetric information. The supplier offers quantity discounts to incentivize retailers to purchase higher values. By adopting blockchain technology, the parties can engage in a group purchasing strategy. Additionally, designing a smart contract, they can automate transactions contingent on realized uncertainties. This setup allows retailers to benefit from discounts across a wider range of orders. However, it requires them to reveal their orders to competitors. We investigate the profitability of supply chain members adopting blockchain technology and smart contracts for group purchasing transactions.

4 - Impact of Average Rating and Number of Reviews on Business Performance on Digital Platform

Samayita Guha, Florida International University, Miami, FL, United States

Consumers increasingly use social media content to make their purchase decisions. Many consumers study online reviews about products or services before purchasing. Online review platforms, such as Yelp and TripAdvisor, provide consumers with reviews and ratings of businesses and service providers. Hence, all types of businesses maintain their presence on online review platforms to improve their performance. To make their presence felt strongly among so many other competitors on online review platforms, businesses need to understand which review characteristic (such as volume, rating value, reviewer identity, texts, etc.) has more influence on consumer purchase decisions and in what ways. In this context, it becomes important to understand the interplay between both review rating and review number. To study this interplay empirically, we use a data set about restaurants on the Yelp platform. The answer to our research question will help restaurants allocate their resources more effectively to improve the number of reviews and ratings.

MD65

Regency - 603

What Should We be Teaching Undergraduates in the Age of ChatGPT?

Panel Session

Education Outreach

Co-Chair: Ken Murphy, Merage School - UC Irvine, 4293 Pereira Drive, Irvine, CA, 92697, United States

1 - Panelist

Patrick Johanns, University of Iowa, Iowa City, IA, United States

2 - Panelist

Shannon McKeen, Wake Forest University, Winston Salem, NC, United States

3 - Panelist

Gustavo Malpele, Expedia Group, Seattle, WA, United States

4 - Panelist

Aaron Burciago, PrimeAI, Leesburg, VA, United States

MD66

Regency - 604

AI-human Interaction and Applications

Invited Session

Artificial Intelligence

Chair: Yuan Zhang, University of Memphis, Memphis, TN, United States

1 - Delving into Generative AI's Role in Mental Healthcare

Polina Durneva, University of Memphis, Memphis, TN, United States, Yuan Zhang

Mental health is a significant public health concern, which imposes a substantial economic burden on the healthcare system. Mental health disorders reduce quality of life and increase risks of disability, morbidity, and mortality. Limited availability and affordability of mental health services, due to shortages of professionals, geographic barriers, and lack of insurance coverage, pose significant challenges. Consequently, exploring alternative avenues for addressing mental health issues is essential. The advent of ChatGPT represents a transformative leap in self-management of mental health. Generative AI technologies like ChatGPT simulate human-like interactions, generate adaptive responses, and tailor conversations based on individual user inputs. However, the specific ways users interact with ChatGPT for mental health purposes remain unclear. Unlike structured mobile health apps, ChatGPT offers greater versatility and adaptability. Understanding user interactions with ChatGPT is crucial for identifying key variables and behaviors that influence desired outcomes, such as improved mental health. This study conducted 20 semi-structured interviews to explore participants' experiences using ChatGPT for mental health self-management. Using grounded theory and constant comparative methods, the analysis identified the process of user interaction with ChatGPT for mental healthcare as well as various use cases. Our findings provide insights into user interactions with ChatGPT, thereby informing the design of future experimental studies and guidelines for health technology developers. This research aims to enhance the accessibility and effectiveness of mental health support, ultimately improving the well-being and quality of life for individuals facing mental health challenges.

2 - Can Voice-Based AI Help US Feel Less Lonely: a Perspective Beyond Anthropomorphism

3 - Mindfulness in AI Tool Usage Among Software Developers: Impact on Innovation Activities

Yafang Li, University of Memphis, Memphis, TN, United States, Yujie Zheng, Pavankumar Mulgund, Raghvendra Singh

Generative artificial intelligence (AI) has become an integral component in the software development landscape, enhancing productivity and efficiency through capabilities like intelligent searching, recommendation systems, and automated code generation. While these tools facilitate tasks such as generating boilerplate code, explaining complex algorithms, and conducting research, there exists a potential risk of over-reliance which may undermine essential human-centric research qualities such as critical thinking and ethical judgment. Thus, mindful use of AI tools is crucial. This study explores the influence of mindfulness in the use of AI tools on innovation activities among software developers. We categorize the mindfulness states of software developers into high, medium, and low, assessing how each impacts the utilization of AI in two primary innovation activities: exploration and exploitation. Exploration involves using AI tools to delve into unfamiliar territories for diversification, while exploitation focuses on leveraging familiar domains to enhance specialization. This research aims to provide insights into how different levels of AI mindfulness affect the innovative outputs in software development environments, contributing to a balanced approach to AI tool integration in creative and technical processes.

4 - Applying Inverse Reinforcement Learning to Understand User Behavior in Online Health Communities

Yaqi Zhang, Central University of Finance and Economics, Beijing, China, People's Republic of, Xi Wang, Zhiya Zuo

This study seeks to understand how receiving social support influences various motivational types and, consequently, different behavioral patterns. Setting out from the theoretical lens of incentive salience, we posit that user motivations dynamically evolve during their initial support seeking upon joining an OHC, hence adapting their engagement behaviors progressively. Specifically, we applied inverse reinforcement learning to identify the "motivations" of diverse engagement patterns of users by inferring the rewards underlying engagement behavior types contingent on the types of social support received by users. Furthermore, we employed the panel autoregression model to examine the mutual influence among users' behavioral patterns and underlying motivations. Our findings provide valuable insights for community administrators and information system designers to maintain and enhance user activity in OHCs.

5 - Incorporating Human Preferences into Interpretable Reinforcement Learning with Tree Policies

Stephanie Milani, Carnegie Mellon University, Pittsburgh, PA, United States

Interpretable reinforcement learning (RL) seeks to create agents that are efficient, transparent, and understandable to the populations that they impact. A significant gap in current approaches is the underutilization of human feedback, which is typically employed only for post-hoc evaluation. We propose to center the needs of end users by incorporating the feedback that would be obtained in a user study directly into the training of interpretable RL algorithms. Our approach involves preference learning, where we learn preferences over high-level features that are not directly optimizable during the RL training process. We introduce an evolutionary algorithm that leverages user feedback to guide training toward interpretable decision-tree policies that are better aligned with preferences. We demonstrate the effectiveness of our method through experiments using synthetic preference data on three RL environments. We show that our approach not only yields policies that are more aligned with underlying user preferences but does so with sample efficiency in the number of user queries, thereby decreasing the burden on the user in providing such data.

MD67

Regency - 605

AI and Digital Platforms

Invited Session

Information Systems

Chair: Lei Wang, University of Utah, Salt Lake City, UT

1 - Standing on the Shoulders of a Genai Platform: Examining the Impact of Implementing Customized Gpts

Yinan Yu, University of Oklahoma, Norman, OK, United States, Hua Ye, Yi-Jen Ho

Generative AI (GenAI) is evolving from a B2C-oriented model to a platform-based ecosystem. While previous studies have focused on GenAI's impact on individual users, the effects on other platform stakeholders remain unexplored. This study investigates the impact of GenAI on business stakeholders, specifically through the adoption of customized GPTs on business developers' websites. Using Look-Ahead Propensity Score Matching (LA-PSM) and Differences-in-Differences (DiD) estimations, our observational study reveals that adopting customized GPTs does not influence traffic volume but significantly enhances visitor engagement metrics. These effects are more pronounced for niche websites and desktop users. A survey of ChatGPT Plus users confirms that customized GPTs provide high-quality and specific information, encouraging deeper user interaction on developers' websites. Our study contributes to the GenAI and two-sided platform literature by highlighting the nuanced effects of GenAI on business stakeholders. Practically, it suggests that businesses can achieve substantial improvements in user engagement, particularly for niche websites and desktop users.

2 - Visualizing the Competitive Market Structure - Mapping Products and Customers Using Online Reviews

Yifan Zhang, Kennesaw State University, Kennesaw, GA, United States, Ning Zhong

Online reviews have become an essential tool for businesses to learn about consumer perception of their brand and products. In this research, we develop a dynamic latent factor model that generates a joint spatial mapping of products and brands to provide insights into the underlying competitive market structure. The proposed model combines review text with review rating through a topic model and simultaneously performs market segmentation and brand positioning. We apply our model to car review data and show how the model may be used by managers to identify changes in consumer perception in online platforms and visualize the competitive market structure.

3 - The Philosopher's Stone for Science - The Catalyst Change of AI for Scientific Creativity

Qian Chen, Penn State University, STATE COLLEGE, PA, United States, Ian Ho, Pin Sun, Dashun Wang

The work investigates whether AI is a catalyst for scientific creativity and what theoretical explanations behind observed AI-supported creativity are. Employing the Logical Creative Thinking (LCT) framework, we conjecture that AI enhances scientific creativity by providing faster search algorithms and offering possibilities to explore new search paths for uncovered knowledge. AI is expected to facilitate creative knowledge *hybridization* (i.e., *recombination* in LCT) across fields and serve as a stimulus for knowledge mutation (i.e., *replacement* in LCT) within a field. Accordingly, we consider two measures of scientific creativity: novelty (as hybridization) and disruption (as mutation). To quantify the AI impact, we analyze the publications from 2000 to 2021 and their citation networks. Our findings first inform that AI increases the novelty of mediocre (medium-level) and the top (90th-percentile) papers while enhancing the disruption of the mediocre papers only. Second, we identify nuanced variations in the impact on creativity across fields. Specifically, AI has the strongest and least impacts on basic science and humanity, respectively. Third, our citation-network analyses further uncover the direct and indirect effects of AI. Citing AI-related papers from other fields fosters novelty due to the hybridization of diverse techniques. Yet, citation concentration within specific or own fields leads to an indirect negative impact on novelty. As for disruption, we observe a similar pattern applied to mediocre papers only. For the most disruptive papers, over-emphasizing specific fields and references hurts scientific mutation, whereas the increase in within-field citations deepens the understanding of prior work and illuminates new reasoning paths.

4 - Understanding the Impact of Algorithmic Viewer Targeting in Live-streaming Commerce

Qinping Lin, Tianjin University, Tianjin, China, People's Republic of, Jing Tian, Lei Wang, Ning Jia

In order to attract viewers with high purchase intention and enhance the viewer-to-buyer conversion rate in the live streaming, many firms are increasingly implementing the algorithmic viewer targeting (AVT) strategy. Based on two unique datasets, we conduct three analyses to explore the effectiveness of AVT strategy. Our study has the following findings: (1) the AVT strategy generally associates with the enhanced interaction atmosphere and increases sales revenue of live streaming. (2) targeted viewers acquired from AVT consume more products compared to organic viewers. (3) AVT strategy enhances the purchase from organic viewers through a spillover effect, facilitating an improved shopping atmosphere with increased commenting interaction and order placement notifications. This study contributes to the emerging live streaming literature by unfolding the impacts of AVT strategy. It also helps firms evaluate their usage of the AVT strategy and optimize their budget allocation accordingly.

MD68

Regency - 606

Advanced Approaches for Network Resilience and Optimization

Contributed Session

Chair: Kedong Chen, Old Dominion University, Norfolk, VA, United States

1 - Restoring a Maximum Flow Network Using Reinforcement Learning

Nayan Chakrabarty, University of Arkansas, Fayetteville, AR, United States, Kelly Sullivan

We examine the problem of restoring network edges that have been disrupted. Specifically, we seek to identify which edges to repair in each time period to maximize the cumulative maximum flow for a given source and sink node. Network restoration problems have been studied from the perspective of designing exact and heuristic solution procedures, but reinforcement learning (RL) algorithms have not been explored extensively for this class of problems. For the deterministic version of the problem, we characterize the necessary conditions for an optimal restoration policy based on structural characteristics of the network. Using this property, we solve the network restoration problem using the maskable proximal policy optimization (PPO) RL algorithm in which masking is used to remove actions that do not meet the necessary conditions. We perform numerical experiments to compare our implementation against benchmark methodologies.

2 - Data-Driven Emergency Medical Service Network Design Models: A Robust Stochastic Dominance Approach

Chun Peng, Beijing Jiaotong University, Beijing, China, People's Republic of, Erick Delage

We study an Emergency Medical Services network design problem and propose two novel data-driven optimization models that account for uncertainty about emergency demand with a time-dependent ambiguity set, leveraging the recent advances on robust optimization. We compare the two models through an out-of-sample analysis using both synthetic and real-life data.

3 - Robust Modeling to Minimize the Total Time of the Process with Random Execution Time and Reduce the Consumption of "Energy Cost" in Parallel Packaging Machines

Samaneh Golara , Northern Illinois's university , Dekalb , IL, United States, Amin Eslampanah , Ziteng Wang

Abstract:

In this research, the problem of minimizing the execution time and the total cost of electricity on a set of parallel batch processing devices where jobs with different sizes are entered dynamically is investigated. To solve this bi-objective problem, an optimization algorithm based on GAMS named "GAMS bi-objective optimization algorithm" is presented. This algorithm is applied using a GAMS programming approach to improve performance in two problem objectives. Based on the delay status of the current batch after the task is added, two candidate lists are generated to limit the search space. In addition, heuristic information is designed for each of the candidate lists to guide the optimal search. Also, targeted local optimization methods are implemented to improve the solution quality. In the end, the performance of the proposed algorithm is compared with the existing GAMS algorithms using comprehensive simulation experiments. The results of experiments show that the proposed algorithm has a significant improvement in performance over other compared algorithms, especially for large problems.

Keywords: Robust modeling, Parallel packaging machines, Random execution time, and energy cost

4 - Improving the Resilience of an Evolving Capacitated Network Using Graph Neural Networks

Seyyed Farid Hashemian, University of Arkansas, Fayetteville, AR, United States, Maryam Aghamohammadghasem, Jose Carlos Hernandez Azucena, Haitao Liao

This research presents a data-driven model that aims to improve the reliability and resilience of a network under stochastic conditions by determining the sequence of operations. To achieve this, an objective function is designed to improve the network's reliability. Surrogate models such as Monte Carlo simulation and Graph Neural Networks (GNN) are utilized to quickly determine the objective function's value based on a sequence of approximate solutions. The GNN model distinguishes itself from previous research by not requiring additional input information like reliability upper bounds or graph features. Instead, the reliability of the network solely depends on data from the network's edges and nodes. However, the data for this model is imbalanced, which presents a challenge, and there is a need to select samples for the prediction model carefully. After computing the objective function using surrogate models, we need to solve the optimization problem to find the best sequence of actions to maximize the reliability and resilience of the network. To do this, heuristic models are developed that are proficient in dealing with complex scenarios and the sequential nature of the design problem. The capability of this type of data-driven model to perform optimization under stochastic settings that are dynamically evolving, as well as the ability to discover extensive decision spaces, led to the decision to use heuristic models.

5 - Clustering in Cardinality-Constrained Portfolio Optimization

Amirmohammad Ebrahimi, University of Florida, Gainesville, FL, United States, Hamed Amini

In portfolio optimization, efficiently managing large pools of assets while adhering to cardinality constraints presents a significant challenge. We propose a novel portfolio optimization framework that combines cardinality constraints with the classical Markowitz mean-variance model, using clustering to reduce dimensionality and achieve an optimal balance of risk and return. We use spectral clustering to group the residual returns of stocks. This method reveals natural groupings of assets based on their returns and correlations and enhances our understanding and categorization of assets, which is crucial for efficiently reducing the optimization space and dimensionality. Our empirical results show that our method is effective in managing portfolios while limiting the number of assets in each group. By optimizing the number of asset clusters, we make the asset selection process easier and achieve performance close to that of an unconstrained portfolio. This demonstrates that clustering strategically can maintain near-optimal performance and provides a robust framework for large-scale portfolio management under real-world constraints. This study shows that our framework improves investment strategies, making portfolio diversification and risk management more efficient and manageable.

6 - Managing Reputational Risks Under Institutional Pressure

Kedong Chen, Rensselaer Polytechnic Institute, Troy, NY, United States

Managing firm reputation risks is imperative as firm reputation impacts its profits, employee and customer loyalty and market value. We posit that supply chain relationships and national culture play important roles on firm reputational risks. We empirically examine the impact of supply network and the moderation role of national culture.

MD69

Regency - 607

Policy and Digital Platforms

Invited Session

eBusiness

Chair: Sameer Borwankar, McGill University, Montreal, QC, Canada

1 - Effects of Corporate Sociopolitical Activism

Sameer Borwankar, McGill University, Montreal, QC, Canada, Shrijata Chattopadhyay

As political polarization deepens across the US, large corporations have a significant role to play in the public discourse regarding political issues. Yet it is uncertain how corporations should react to social issues that have been politicized over the years. We study the politicization of LGBTQ+ rights over the years in the U.S. and its impact on corporate stances on socio-political issues. As corporations go beyond profit maximization and take stances on socio-political issues, it is difficult to satisfy all stakeholders and hence, can result in reduction in firm value. On the other hand, strategic use of socio-political stances by corporations can enhance firm value. We find a decline in public stances by corporations post-politicization of LGBTQ+ rights, although the stock market has a heightened positive reaction to such stances post-politicization.

2 - When Moderators are Moderated: Evidence from an Online Forum

Ilango Guru Muniasamy, Daniels School of Business, Purdue University, West Lafayette, IN, United States, Wreetabrata Kar, Hossein Ghasemkhani, Warut Khern-am-nua

Online communities play a vital role in creating safe and inclusive digital spaces based on shared interests. The success of these communities hinges on cultivating trust and a sense of belonging, with community moderation and disciplinary processes playing a crucial role in building this trust. While research has explored the impact of disciplinary actions like user suspensions on subsequent content contributions, the influence of suspensions on participation in community moderation remains unexplored. Our study fills this gap by examining how suspensions affect users' engagement in upvoting and downvoting, key tools for moderation of online communities.

Utilizing data from a leading Asian online forum, we discover that post-suspension, users tend to downvote more frequently but exhibit no change in upvoting. This behavior is accompanied by a short-term decline in the quality of downvoting and conformity with community consensus, revealing evidence of an antisocial response to suspension. This response is amplified among the set of core users who contribute more content and are longer tenured in the forum, raising additional concerns about the negative effects of suspension.

To interpret these behaviors, we draw on the Multi-motive Model of Reactions to Interpersonal Rejection - a prominent theory that studies individuals' response to rejection, and the Temporal Need-Threat model which evaluates individuals' evolution of the responses over time. Our findings suggest that suspensions, though intended as corrective measures, may inadvertently disrupt platform harmony in the short-term by promoting hostile moderation activities

3 - The Impact of Online City Disclosure

Tiange Li, McGill University, Montreal, QC, Canada, Animesh Animesh, Sameer Borwankar

Studies suggest online communities can mitigate spatial inequality across regions of different development levels and facilitate more equitable discussion. Yet, this digital egalitarianism is at risk in the face of a pervasive trend toward increased transparency, particularly when platforms disclose user location data. Such disclosures may unintentionally exacerbate regional disparities by reinforcing pre-existing tensions and biases. In this study, we aim to fill this gap and investigate how city disclosure influences the network significance of cities in terms of online interactions. Specifically, we investigate how users in higher-tier cities respond to city disclosure (i.e., engage and interact) compared to their lower-tier cities counterparts, in the context of the Chinese city-tier system. By leveraging a natural experiment on China's civil service exam forum, we show that city disclosure introduces location-induced social hierarchy bias into online interactions: users located in higher-tier cities are less likely to engage with or receive responses from others, demonstrated by a decrease in both outdegree and indegree centralities of higher-tier cities. However, our analysis suggests that certain real-world competencies, such as educational attainment or local competitive dynamics, may mitigate this bias. Our robustness checks, including provincial data limitations, region subsample analysis and alternative social hierarchy indicators, confirm the consistency of our results. Additionally, we enrich our study with a dyad-level analysis and show that city disclosure significantly increases same-status competition, be it within the same tier group, tier, or province.

4 - Monetization and Accreditation in Healthcare

Krishna Pothugunta, Michigan State University, East Lansing, MI, United States, Xiao Liu, Anjana Susarla, Rema Padman

Monetization and accreditation in healthcare, particularly on social media with the rapid growth of user generated content, presents a unique opportunity and challenge to maintaining inclusivity on the social media. Since the introduction of YouTube Partner Program in 2007, the program has gone through various changes, notably through 2017 to 2023. Prior literature has shown that monetization models like subscription services or premium content can create access barriers, thereby affecting the representativeness of the platform by excluding certain demographics or socio – economic groups. In May 2020, YouTube announced that all healthcare professionals must disclose their credentials in videos related to health or medical topics. Even though the aim of this policy is to enhance the transparency and credibility of the content creator or presenter, but stringent criteria may exclude valuable contributors who lack formal credentials, thereby negatively affect the representativeness of the platform by limiting the diversity. So, both monetization and accreditation can impact the representativeness of healthcare in social media platforms. The study intends to measure the change in representativeness in the videos from 2018 to 2023 by using multiple channels related to healthcare. The major focus of the study is towards understanding the major difference in representativeness in videos that belong to channels that are accredited and influencers on YouTube, by using a difference – in – difference method. The contribution of the study is to address the change in level of inclusivity and representativeness on social media with the respective modifications in monetization and accreditation policies.

MD70

Regency - 701

Leveraging Machine Learning for Advanced Demand-Side Electricity Management

Invited Session

ENRE: Electricity

Chair: Lingxiang Yun, Taxes A&M University Corpus Christi, Corpus Christi, United States

1 - Sample-Efficient Deep Learning for Demand Response

Guangchun (Grant) Ruan, Massachusetts Institute of Technology, Boston, MA, United States

The superior performance of deep learning relies heavily on a large collection of sample data, but the data insufficiency problem turns out to be relatively common in global electricity industry. How to prevent overfitting in this case becomes a fundamental challenge when training deep learning models in different demand response applications. With this in mind, we propose a general framework, namely Knowledge-Augmented Training (KAT), to improve the sample efficiency, and the main idea is to incorporate domain knowledge into the training procedures of deep learning models. Specifically, we propose a novel data augmentation technique to generate some synthetic data, which are later processed by an improved training strategy. This KAT methodology follows and realizes the idea of combining analytical and deep learning models together. Modern learning theories demonstrate the effectiveness of our method in terms of effective prediction error feedbacks, a reliable loss function, and rich gradient noises. At last, we study two popular applications in demand response decision making: user modeling and probabilistic price forecasting. The proposed method outperforms other competitors in all numerical tests, and the underlying reasons are explained by further statistical and visualization results.

2 - Data-Driven Robust Unit Commitment with Statistical Guarantees: An Integrated Forecasting and Optimization Approach

Yue Chen, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Rui Xie

The penetration of distributed renewable generation increases the demand side randomness and brings new challenges for unit commitment. Following the idea of predict-and-optimize, this paper proposes a novel robust unit commitment approach that integrates the forecasting and optimization processes. In the forecasting part, besides directly using local load data, horizontal federated learning is adopted to utilize the similarity between loads, and a privacy-preserving vertical federated learning framework is developed to exploit consumers' private data. These three kinds of predictions are combined by a weighted summation, where the weight is adjusted based on the test performance of the optimized unit commitment strategy. In the robust unit commitment part, the uncertainty set is constructed in a data-driven manner based on the combined prediction and the historical forecast error so that it has a statistical guarantee. It is then reconstructed using a shape derived from the robust unit commitment problem to mitigate conservativeness. The problem is formed as two-stage robust optimization and solved by the C&CG algorithm. Case studies on the modified IEEE 30-bus and 118-bus systems demonstrate the effectiveness of the proposed approach.

3 - Enhanced Load Detection with Data-Driven Appliance Signatures Using Mixed Integer Quadratic Programming in Non-Intrusive Load Monitoring

Jay Rosenberger, University of Texas at Arlington, Arlington, TX, United States, Marina Materikina, Victoria Chen, Wei-Jen Lee

Rapidly developing renewable energy sources pose new challenges for utility companies' production and demand response (DR) planning. Studying customers' electricity consumption behavior can enhance the effectiveness of DR programs. Non-intrusive load monitoring (NILM) enables the analysis of individual appliance electricity patterns, fostering the connection between customers and energy providers. A mixed integer quadratic programming algorithm utilizing data-driven appliance signatures was developed for NILM. The algorithm has the ability to study very complex electrical devices, detect their signal with higher accuracy and reconstruct the electricity usage over time. The detailed signatures for modern state-based appliances (EV, AC, solar, washer, dryer) were studied using nonlinear programming and piecewise functions that depend on mode shapes. This approach significantly improved load detection accuracy during the disaggregation phase, and was applied to low-frequency data making NILM process more practical.

4 - Reinforcement Learning for Dynamic and Heterogeneous Incentive-Based Demand Response

Elson Cibaku, New Jersey Institute of Technology, Rutherford, NJ, United States, Bahareh Kargar, SangWoo Park

This paper introduces a Reinforcement Learning (RL) approach for dynamic incentive-based demand response (DR) programs, designed to maximize the operating profit for load aggregators. Our methodology integrates a Markov Decision Process (MDP) model with a Deep Q-Network (DQN)-based RL framework, allowing for sophisticated decision-making that adapts to changing consumer behavior patterns and system states. By simulating interactions between an aggregator and multiple heterogeneous customers, our model dynamically adjusts the optimal incentive pricings at both an individual and appliance-specific level. Additionally, our model incorporates a comprehensive reward function that includes multiple factors such as real-time locational marginal price (LMP), cost of self-owned generators, and cost of battery operations, enhancing the realism of modern load aggregators. We validate our approach through extensive simulations and a real-world dataset, demonstrating its potential to effectively balance energy supply and demand by adapting to real-time market conditions and consumer preferences.

5 - Decentralized DER Integration into Wholesale Energy Markets via Mean-Field Games

Jun He, Purdue University, West Lafayette, IN, United States, Andrew Liu

FERC Order 2222 facilitates the integration of distributed energy resources (DERs) into wholesale energy markets, but specific mechanisms are needed for effective small prosumer participation. A key concern is prosumers' hesitation to relinquish control of their assets to aggregators. Current research primarily focuses on how aggregators can bid into a wholesale market and design contracts for direct load control of the DERs. In response, we propose a decentralized model that allows prosumers, especially those with solar PV panels and energy storage, to make autonomous decisions based on real-time locational marginal prices (LMPs). Our solution employs a mean-field-based algorithm suited for a multi-agent system with a large number of agents. This algorithm enables prosumers to maintain beliefs about LMPs at a long-run equilibrium, solve their own Markov decision problems, and determine their supply or demand bids, which are then aggregated for the wholesale market. System operators continue running multi-settlement systems, including unit commitment and economic dispatch. We prove the existence of a mean-field equilibrium (MFE) and provide a reinforcement-learning-based algorithm for prosumers to learn to converge to an MFE. Numerical results show that LMPs can quickly reach a steady state. Furthermore, numerical comparisons with and without energy storage illustrate that our model can prevent extreme LMP values, fostering a more stable market even with completely decentralized decision-making.

MD71

Regency - 702

Advances in Data Science: From Description to Prediction and Decision-making

Invited Session

Data Mining

Chair: Naichen Shi, University of Michigan, Ann Arbor, MI, United States

1 - Robust Personalized Design via Bayesian Optimization Under Covariate Uncertainty

Jiacheng Liu, University of Michigan, Ann Arbor, MI, United States, Wenbo Sun, Jingwen Hu, Jionghua Jin

Personalized design is to optimize a functional design policy that assigns the optimal design based on human covariates. Despite the noiseless setting in the computer simulations, applying the optimal design presents reliability concerns due to the potential differences between the selected and true optimal designs. The presence of measurement errors in covariates can further lead to discrepancies in the system response performance. The high computational cost of data collection also limits the ability to build an accurate surrogate model for the entire population. To address these challenges, we propose a robust Bayesian Optimization framework to minimize the discrepancy in system response by imposing penalties on the variation of system responses induced by covariate uncertainties. An appropriate acquisition function is introduced to consider the mixed type of categorical and continuous design variables when searching for new samples. This robust design policy reduces the sensitivity of the optimal design policy against covariate uncertainties as well as maintains system performance. Both the numerical examples based on the synthetic data and vehicle restraint system design demonstrate the efficiency and robustness of the proposed method.

2 - Robust Collaborative Online Learning Under Adversarial Attack

Tanapol Kosolwattana, University of Houston, Houston, TX, United States, Huazheng Wang, Ying Lin

Dealing with multiple different but closely related sequential decision-making problems is usually achieved by the collaborative online learning algorithms. However, many decision-making problems, such as pay-per-click advertisements and recommendation systems, face risk of adversarial attacks, leading to vulnerabilities in existing collaborative online learning algorithms. To the best of our knowledge, there is no robust online learning algorithms developed for a population of dependent sequential decision making under adversarial attack. To mitigate this gap, this study proposes a robust collaborative online learning method that captures latent representative models inherent in the population through representation learning. Additionally, we design a novel robust collaborative UCB algorithm to estimate representative models from sequentially observed data corrupted by adversarial attacks. We demonstrate the effectiveness of our method through theoretical analysis and simulation studies.

3 - Online Optimal Testing Policy Learning

Qiyuan Chen, University of Michigan, Ann Arbor, MI, United States, Raed Al Kontar

Experimentation is one of the most fundamental tools for scientific discovery and operations research for understanding the unknown nature. In many real-world applications like health examination and quality control, a sequence of experiments is often needed to probe the latent state of an object of interest. However, the number of experiments required is often much less than the number of experiments available. For example, a doctor usually does not recommend doing all the medical tests in a hospital, although doing all of them will be sufficient to diagnose a patient. Depending on the results of the previous test, the doctor shall determine the best test to conduct next so that the patient can be diagnosed with the fewest tests possible. However, the goal of saving testing costs poses fundamental challenges to the learning of this optimal testing policy. There are two consequences of not conducting all the tests. First, one will never know the exact quality of a part, meaning the reward associated with the conducted action is not revealed. Second, the missing values in the historical datasets make the empirical estimators heavily biased.

4 - A tractable approach to solve misspecified Predict-then-Optimize Problems

Omar Bennouna, MIT, Cambridge, MA, United States, Jiawei Zhang, Saurabh Amin, Asuman Ozdaglar

We study the predict-then-optimize framework approach, a common application of machine learning, which entails estimating unknown parameters of a linear optimization problem and then solving the optimization task with these estimations. For example, consider an energy allocation problem when the energy cost in different areas is uncertain. Despite the absence of precise energy cost values at the time of problem-solving, machine learning models are employed to predict these costs, and the resulting optimization problem, which consists for example of minimizing energy costs while meeting some minimal requirements, is solved using state-of-the-art optimization algorithms. We focus on the unrealizable setting, i.e. when the used hypothesis set (predictor function class) does not contain the ground truth value. In this case, there is no known algorithm that solves this predict-then-optimize problem. We provide a tractable algorithm which successfully finds an optimal solution in this setting.

5 - Multi-physics Guided Diffusion Models with Manufacturing Applications

Naichen Shi, University of Michigan, Ann Arbor, MI, United States, Hao Yan, Raed Al Kontar

We present a generic physics-informed generative model that integrates multi-fidelity physics simulations with diffusion processes. Our model categorizes multi-fidelity physics simulations into inexpensive and expensive simulations, depending on the computational costs. The inexpensive simulations, which can be obtained with low latency, directly inject contextual information into diffusion models. Furthermore, when results from expensive simulations are available, we refine the quality of generated samples via a guided diffusion process. This design separates the training of a denoising diffusion model from physics conditional probability models, thus lending flexibility to practitioners. Our model builds on Bayesian probabilistic models, which provide a convenient approach for uncertainty quantification in prediction. We use a numerical simulation in fluid dynamics and a case study in laser-based metal powder deposition additive manufacturing to demonstrate how our model seamlessly integrates multi-fidelity physics simulations and observations to obtain superior predictive performance.

MD72

Regency - 703

Applications of Data Mining Techniques in Biomedical Data Analysis

Invited Session

Data Mining

Chair: Abhidnya Patharkar, ARIZONA STATE UNIVERSITY, Chandler, 85286, United States

Co-Chair: You Zhou, Arizona State University, Tempe

1 - EE-SFE: An Entropy-based Feature Extraction Framework on Subgraphs for Cognitive Disease Classification

Jiajing Huang, Arizona State University, Tempe, AZ, United States, Yiming Che, Teresa Wu

Brain networks play an essential role in cognitive-related disease studies, providing a powerful framework to capture the functional connectivity between different regions of the brain. While existing methods demonstrate prevailing computational efficiency, their lack of transparency hinders the interpretability of the extracted features. Therefore, efficient extraction of meaningful features on these networks for effective brain disease diagnosis is critical yet challenging. To address the challenges outlined above, we introduce a novel two-staged framework named Eigen Entropy-based Subgraph Feature Extraction (EE-SFE) designed for fMRI data. In the initial stage, significant edges, representing pairwise connections between brain regions that exhibit discriminatory power between dysfunctional and healthy groups, are identified through hypothesis testing. Subsequently, these edges are organized into clusters, forming distinct subgraphs. In the second stage, features are extracted from these subgraphs utilizing the Eigen-Entropy (EE) method, a technique previously developed for classification purposes. Our proposed method demonstrates better performances compared to a heuristic approach reported in the literature. More importantly, features extracted by our method retain meaningful interpretation, aligning consistently with findings in neuroscience literature.

2 - Smart Driving: Using Machine Learning for Non-Invasive Detection of Age-Related Cognitive Decline

Firas Al-Hindawi, Arizona State University, Tempe, AZ, United States, Teresa Wu, Erica Forzani, Peter Serhan , Francis Tsow, Yonas Geda

The rise of machine learning techniques has paved the way for innovative approaches in monitoring human health. In this study, we explore the utilization of multi-sensing devices installed in subjects' vehicles to detect age-related cognitive decline. Our focus lies in leveraging various data streams from accelerometer, gyroscope, and GPS sensors to find patterns indicative of cognitive health. We conducted experiments consisting of 20 subjects, including 12 healthy individuals and 8 diagnosed with Mild Cognitive Impairment (MCI). These subjects underwent two distinct sets of experiments: standardized and naturalistic driving scenarios. In the standardized experiments, subjects were instructed to perform specific driving maneuvers within a controlled environment. This controlled setting allowed for the systematic collection of data under various driving conditions. Conversely, the naturalistic driving experiments provided insights into subjects' driving behaviors during their routine travel patterns. By observing their driving habits in real-world scenarios, we aimed to capture a holistic understanding of their cognitive performance during everyday tasks. Through the application of machine learning algorithms to the collected data, we sought to identify subtle deviations in driving behavior that may signal cognitive decline. Our findings demonstrate the potential of non-invasive monitoring techniques in early detection and intervention for age-related cognitive impairments and the advancement of future research in cognitive health assessment through smart driving technologies.

3 - TP-CL: A Novel Temporal Proximity Contrastive Learning Approach for Obstructive Sleep Apnea Detection using Single-Lead Electrocardiograms

Fulin Cai, Arizona State University, Tempe, AZ, United States

Early detection of obstructive sleep apnea (OSA) benefits individuals by improving overall life quality and preventing malignant complications associated with untreated OSA. Deep learning-based applications using single-lead electrocardiograms (ECG) have shown promise in automatic OSA detection. However, most efforts have not fully considered variations in the detected signals of breathing events. A breathing event is highly conditioned to the present sleep context (sleep position and stage), while sleep context varies through time. In addition, the high variance among apnea events, associated with event duration, recovery period after apnea, and respiration pattern, shall be utilized to support OSA detection. Utilizing these two characteristics, a temporal proximity contrastive learning (TP-CL) approach is proposed for OSA detection. TP-CL has three modules: (1) a representation learning module using a DL-based encoder; (2) a temporal delta classification module using the difference between temporal proximities to support OSA detection; and (3) a temporal proximity contrast module employing a novel contrastive-based loss function to form a representation space with minimal contextual influence. Experimental results show that TP-CL can achieve competitive results, outperforming the state-of-the-art OSA detection models, in both subject-dependent and subject-independent settings. A sensitivity analysis showcases the robustness of the proposed approach and provides insight into potential future applications of TP-CL. In addition, using 1-minute ECG signals as input, the proposed TP-CL method provides real-time detection capability and is a potential tool for early OSA detection and in-home OSA monitoring.

MD73

Regency - 704

Learning Control and Decision-Making of Dynamic Systems

Invited Session

Data Mining

Chair: Shuo Wang, University of Texas at Arlington, Arlington, TX, United States

1 - Learning Control of a Complex Population System

Shuo Wang, University of Texas at Arlington, Arlington, TX, United States

Designing a robust control to steer the collective behavior of ensemble systems has been of significant importance and practical demand in engineering and science, such as quantum control, cancer treatment, neurons' spiking patterns, etc. However, such a robust control may be too restrictive and costly to realize focusing on the entire population of the targeted dynamic systems. Instead, a relaxed control synthesis that aims at controlling only the majority of the population would be more practically appealing. In this work, an optimal control strategy is developed and explored in such a sense, that a partial differential equation is used to describe the collective dynamic behavior of the targeted systems, and machine learning techniques are adopted to be combined with the standard necessary conditions from optimal control theory such as Hamilton-Jacobi- equations. Several numerical applications are included to demonstrate the applicability of this proposed method.

2 - Surrogate model of dynamical systems in effective region

Wei Kang, Naval Postgraduate School, Monterey, CA, United States

In this study, we develop learning-based surrogate model of dynamical systems for efficient state estimation and prediction in a limited area. Our model employs a feedforward neural network for online computation, eliminating the need for integrating high-dimensional limited-area models. This approach offers significant computational advantages over integrating high dimensional system models. Furthermore, our method avoids the requirement of lateral boundary conditions for the limited area model in both online and offline computations. The design of our surrogate model is built upon a robust theoretical framework that leverages two fundamental concepts: observability and effective region. The concept of observability enables us to quantitatively determine the optimal amount of observation data necessary to achieve accurate approximation. Meanwhile, the concept of effective region substantially reduces the computational burden associated with computing observability and generating training data.

3 - Research on the Evaluation of Technological Innovation Capability of Intelligent Manufacturing Enterprises

Mengshan Zhu, Tongji University, Shanghai, China, People's Republic of, Chunyan Duan, Wenyong Zhou

The emerging models of intelligent manufacturing bring new opportunities for enterprise innovation. However, innovation inherently is a high-risk research and development activity. To mitigate losses and transform threats into opportunities, intelligent manufacturing enterprises must effectively control innovation risks. This paper begins by identifying innovation risks and analyzing their sources, constructing a comprehensive evaluation index system that includes dimensions such as technological risk, market risk, and management risk. Subsequently, the multiple intelligent algorithms-based innovation risk evaluation model is developed and applied in a case study. Furthermore, by comparing artificial intelligence evaluation models such as Support Vector Machine (SVM) and Decision Tree (DT), the superiority of the proposed method is demonstrated, showing optimal performance in evaluating innovation risks in intelligent manufacturing enterprises. Finally, relevant risk management measures and recommendations are proposed, aiming to enhance the innovation capacity and market competitiveness of intelligent manufacturing enterprises.

4 - Decision supporting for car sharing platform selection based on risk preferences and heterogeneous uncertain data

Hui Lin, School of Economics and Management, Tongji University, Shanghai, China, China, People's Republic of, Xueyang Zhang, Jianxin You

The choice of car sharing platforms (CSPs) is crucial for users to obtain a high-quality, safe and environmentally friendly car experience. Usually, this selection process involves various CSPs and attributes, which can be regarded as a heterogeneous multi-attribute group decision making (MAGDM) problem. Two methods are proposed for considering risk preferences and heterogeneous uncertain data, as well as two decision scenarios where the decision information is in forms of numerical-linguistic hybrid and linguistic hybrid types. Firstly, to capture the impact of individual risk preferences (IRPs) on multi-granular interval linguistic information in heterogeneous decision making, a semantic representation model and a transformation model of multi-granular interval linguistic terms under IRPs are defined. Secondly, in heterogeneous MAGDM with incompletely known weight information under different IRPs, we formulate some optimization models for solving weights of attributes and decision makers separately. These models are grounded on the deviation maximization and the group compatibility principles. Based on the aforementioned models, an approach to addressing heterogeneous MAGDM problems that involve incomplete weight information and IRPs is presented. Thirdly, in heterogeneous linguistic MAGDM with completely unknown individual attribute weights under IRPs, the best-worst method (BWM) is extended to the case of different individual attribute sets. An optimization model for solving attribute weights in individual attribute sets is constructed. Based on proposed models, a method for heterogeneous linguistic MAGDM problems under diverse IRPs is proposed. Finally, an example of car sharing platform selection and comparative analysis are provided to clarify the effectiveness and feasibility of our proposed methods.

5 - Data asset quality evaluation Framework based on a Hybrid Multi-criteria Decision-making Method

Tao Xu, Tongji University, Shanghai, China, People's Republic of, Jianxin You, Luning Shao, Mengmeng Shan

This paper presents a new framework for evaluating data asset quality using a hybrid Multi-Criteria Decision-Making (MCDM) approach that integrates the Decision Making Trial and Evaluation Laboratory (DEMATEL), Best-Worst Method (BWM), and Fuzzy- Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) techniques. Firstly, the framework consider data as both a product and an asset, leading to the quality indicators being developed beyond the traditional dimensions. Subsequently, the interrelationships among indicators are addressed using the DEMATEL method, allowing for the identification of key indicators that significantly influence data asset quality in a given scenario. The BWM method is then employed to determine the weights of these key indicators, enabling a more precise assessment of their importance. After that, the TOPSIS method, incorporating triangular fuzzy numbers, is utilized to rank the data asset quality of different companies. Finally, the effectiveness of the framework is demonstrated by applying it to a group of companies, and the results of the company's evaluation are discussed, along with the corresponding data asset quality improvement initiatives.

MD74

Regency - 705

Optimization Methods for Grid Operations and Market Clearing

Invited Session

ENRE: Electricity

Chair: Feng Qiu, Argonne National Laboratory, Lemont, IL, United States

1 - Can a Storage System Exercise Market Power in an Electricity Market?

Feng Tu, Ohio State University, Columbus, OH, United States, Antonio Conejo

We analyze the ability of a storage system to exercise market power within an electric energy system. For this, we use three models. The first one represents the storage system as a price taker, the second one as a price maker, and the third one allows the system operator to freely operate the storage system for the best of the power system as a whole (maximum social welfare). Considering analytical outcomes from a stylized model and numerical simulations from a realistic model, we found that the storage system has no ability to exercise market power. Thus, the outcomes of the price-maker model and those of the model in which the system operator is in control of the storage system are equal. In other words, the objective of the storage system as a price maker -- maximum own profit -- and the objective of the system operator -- maximum social welfare -- are aligned, and thus, a bi-level model is not needed. Finally, we note that solving the model in which the system operator is in control of the storage system is simple, but solving the price-maker model, which is bi-level, is complex.

2 - Congestion relief through market-to-market coordination

Feng Qiu, Argonne National Laboratory, Lemont, IL, United States

The growing usage of renewable energy resources has introduced significant uncertainties in energy generation, posing challenges for Regional Transmission Operators (RTOs) in managing transmission congestion. The cost of real-time congestion for the Midcontinent Independent System Operator (MISO), for example, surged to \$3.7 billion in 2022, more than tripling since 2020. To mitigate congestion that affects neighboring regions, RTOs employ a market-to-market (M2M) process, in which they exchange real-time security-constrained economic dispatch solutions and communicate requests for congestion relief. While this method provides some economic benefits, it still struggles with issues like power swings and time delays. In this presentation, we analyze the M2M process to better understand its efficacy and identify potential improvements. We develop an open-source version of the M2M process using UnitCommitment.jl and test it across a comprehensive set of multi-area benchmark instances. To evaluate the potential impact of M2M enhancements, we develop a centralized model that provides a lower bound on congestion costs and compare it against current practice. Additionally, we identify scenarios where the existing M2M process fails to provide solutions. Finally, we explore enhancements and alternative strategies to narrow the efficiency gap in multi-area congestion management.

3 - Solving the security constrained AC unit commitment problem using an enriched Benders decomposition approach

Jorge Orrego, Ohio State University, Clumbus, OH, United States

We address the security-constrained AC unit commitment problem in a corrective manner, considering that only the transmission line fails. We convexify the AC power flow equations using a conic relaxation to achieve tractability. We propose an enriched Bender-type decomposition algorithm to solve large-scale systems, assuming all transmission lines could fail. This algorithm involves three optimization problems: a master problem that provides scheduling decisions, a normal-operation condition problem that dispatches generators, and subproblems that provide operating decisions under contingency conditions. The master problem is a mixed integer linear problem containing normal-operating conditions constraints and a small subset of under-contingency operations. Fixing scheduling decisions from the master problem transforms the normal-operation condition problem and the subproblems into continuous conic optimization problems. In addition, fixing dispatch decisions from the normal-operation condition allows solving the subproblems independently, each corresponding to an under-contingency condition. The master problem is a relaxed version of the original problem, thus providing a valid algorithm lower bound, while the subproblems allow deriving an upper bound. The solution of the subproblems enriches the master problem and the normal-operation condition using (i) Benders' cuts, (ii) incorporating a subset of under-contingency operation in the master and normal-operation condition problem, and (iii) enforcing only congested transmission lines. We solve the master problem, the normal-operation condition problem, and the subproblems iteratively until a convergence condition is satisfied. We illustrate the effectiveness of the proposed algorithm using the Illinois 200-bus system and different demand levels.

4 - Optimal Battery Utilization and Swapping Service Level through Balanced Battery Inventory and Replenishment for an Electric Vehicle Battery Swapping Station Network

Reza Alizadeh, Magna International, Boston, MA, United States

Establishing an ample charging infrastructure is pivotal for the successful deployment of electric vehicles (EVs), mitigating concerns associated with restricted range and extended charging durations. Battery exchange stations arise as a promising remedy to alleviate these waiting periods. These stations act as intermediaries between the power grid and consumers, presenting a rapid alternative to recharging batteries. Ensuring their efficacy and profitability necessitates a robust business and operational model. This endeavor introduces an optimization framework for the operational structure of battery swapping stations, emphasizing balanced battery inventory management. The framework offers features like flexible battery restocking from depots and equitable battery inventory distribution among charging stations. The battery float percentage and battery utilization rate are augmented, eliminating unmet demand and minimizing the overall shipment cost. The findings demonstrate the feasibility of the proposed model as a business case, as well as its efficacy in delivering the necessary service.

MD75

Regency - 706

Biodiversity Conservation

Invited Session

ENRE: Environment and Sustainability

Chair: Jue Wang, Smith School of Business, Queen's University, Kingston, ON, Canada

1 - Saving Whales with Optimal Control

Yu Gong, Smith School of Business, Queen's University, Kingston, ON, Canada, Jue Wang

Collision with large vessel is the leading cause of whale mortality. The collision risk can be reduced by the establishment of vessel speed reduction (VSR) zone, in which large vessels are asked to reduce speed. However, reducing speed is costly to the mariners by causing shipping delays. Thus, the compliance rate is low in VSR. We develop an optimal control model to optimize the VSR locations based on daily prediction of whales' presence distribution. Our model utilizes the dynamic discrete choice model to capture mariners' strategic response. The goal is to minimize the probability of whale mortality subject to a constraint on the economic impact to the shipping industry.

2 - Site Prioritization for Conservation Planning

Quanmeng Wang, Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Guodong Lyu, Chung-Piaw Teo

We propose a new approach to prioritize activities in a resource constrained environment, where the resource budget may change across time. This problem arises frequently in spatial conservation planning, a problem in biodiversity protection worldwide. Traditional approach often adopts greedy techniques to develop a priority ranking of all the zones in the conservation site, without considering resource available for execution. We address this gap and provide a systematic approach in this paper to synthesize the impact of resource uncertainty on the optimal prioritization strategy. We use a well-known real life case study in this field to show how our approach can produce a more resilient plan to cushion the impact of budget changes, in realistic size instances with thousands over zones. More interestingly, the prioritization approach generally produces solution that performs close to the scenario-wise optimal solution at each time period, demonstrating a desirable feature of our approach in practice.

3 - Birdbnb: Data-Driven Pitstop Support for Migrating Birds

Li Chen, Cornell University, Ithaca, NY, United States, Jue Wang, Xiaoyue Yan

Wetland stops serve as vital rest and refueling sites for millions of migrating birds. Many of these wetlands, however, have been lost to farmland conversion over the past, posing a serious challenge to bird populations. The BirdReturns program, a biodiversity initiative, runs an open bidding process to acquire temporary wetland habitats from rice farmers during the critical bird migrating seasons. In this talk, we present an optimal data-driven auction mechanism for the program.

MD76

Regency - 707

Industrial Decarbonization I: Carbon Capture and Hydrogen

Invited Session

ENRE: Other Energy

Chair: Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, 78712, United States

1 - Optimal Subsidies for Carbon Capture Storage

Connor Colombe, University of Texas at Austin, Austin, TX, United States, Benjamin Leibowicz

Carbon capture, utilization, and storage (CCUS), has been widely recognized for its important role in decarbonization. Emerging technologies like CCUS are often adopted slowly until the technology matures and the implementation costs reduce enough for firms to profit from them. Government planners can induce earlier adoption of new technologies through direct and indirect subsidies. In this paper we use a stylized Stackelberg game framework involving a government planner and a CO_2 emitting firm in order to study the impacts of different subsidy levels on CO_2 levels, firm profit, and social-welfare. We analytically derive optimal CCUS subsidy levels under a variety of assumptions and apply our findings to a case study of a coal-fired power plant.

2 - Opportunities and Constraints of Hydrogen Energy Storage Systems

Jacqueline Dowling, Carnegie Institution for Science, Stanford, CA, United States, Tyler Ruggles, Edgar Virguez, Natasha Reich, Zachary Ifkovits, Steven Davis, Anna Li, Kathleen Kennedy, Katherine Rinaldi, Lei Duan, Ken Caldeira, Nathan Lewis

In contrast to battery storage, power-to-hydrogen-to-power (P-H2-P) storage systems provide opportunities to separately optimize the costs and efficiency of the system's charging, storage, and discharging components. The value of capital cost reduction relative to round-trip efficiency improvements of P-H2-P systems is not well understood in electricity systems with abundant curtailed power. We used a macro-energy model to evaluate the sensitivity of system costs to techno-economic characteristics of P-H2-P systems in stylized wind-solar-battery electricity systems with restricted natural gas generation. Assuming current costs and round-trip P-H2-P efficiencies, least-cost wind and solar electricity systems had large amounts of excess variable generation capacity. These systems included P-H2-P in the least-cost solution, despite its low round-trip efficiency and relatively high P-H2-P power discharge costs. These electricity system costs were not highly sensitive to the efficient use of otherwise-curtailed power, but were sensitive to the capital cost of the P-H2-P discharge component. If capital costs of charging and discharging components were decreased relative to generation costs, curtailment would decrease, and electricity system costs would become increasingly sensitive to improvements in the P-H2-P round-trip efficiency. These results suggest that capital cost reductions, especially in the discharge component, should be a key priority for innovation in P-H2-P systems for applications in electricity systems dominated by wind and solar generation. Analysis of underground salt cavern storage constraints in U.S.-based wind and solar scenarios suggests that ample hydrogen storage capacity could be obtained by repurposing the depleted natural gas reservoirs that are currently used for seasonal natural gas storage.

3 - Coordination Problems and Incentive Pass-Through in Carbon Capture, Utilization, and Storage Development

Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, United States, Abdullah Albeladi

Carbon capture, utilization, and storage (CCUS) technologies figure prominently in many proposed pathways for achieving ambitious, economy-wide decarbonization goals. While numerous governments offer subsidies for CCUS, the buildout of CCUS infrastructure has been sluggish to date. In this study, we address two research questions related to policy incentives for CCUS development. First, to what extent do coordination problems among multiple players in the CCUS supply chain hinder CCUS development? Second, how much of the total value of CO_2 capture incentives will pass through to other players downstream in the supply chain? To answer these questions, we develop a game-theoretic model of a CCUS economy formulated as a mixed complementarity problem. It incorporates CO_2 capture subsidies with different levels for geologic storage and utilization applications, in line with the 45Q policy in the United States. We first prove two theoretical results, then apply the model to a case study of the Texas-Louisiana Gulf Coast. On coordination problems, we find that they reduce the total volume of CO_2 capture substantially when the incentive for geologically stored CO_2 is low (by 33% for $50/ton), but have little impact when the incentive is high. On incentive pass-through, we find that CO_2 capturers appropriate the largest share of the subsidies, but greater fractions of the incentives pass through to pipeline operators as the incentives become more generous and CCUS activity expands.

4 - Hydrogen and carbon capture and storage enable the decarbonization of hard-to-abate industries

Alissa Ganter, ETH Zurich, Zurich, Switzerland, Paula Baumann, Veis Karbassi, Giovanni Sansavini

The transition to net-zero emissions proves challenging for industries that inherently rely on carbonaceous feedstocks, such as the chemical industry, steel, cement, and refineries. These hard-to-abate industries are difficult to decarbonize as direct electrification is limited, and cost-efficient decarbonization alternatives are lacking. In this context, the use of low-carbon hydrogen and the coupling with carbon capture and storage emerge as promising solutions that can enable the transition. In addition, biomass feedstocks could provide a cost-efficient low-carbon alternative to fossil fuels. We formulate a linear optimization problem to investigate cost-optimal decarbonization pathways for these hard-to-abate industries in Europe. We consider a multi-year time horizon and use a regional resolution to model the transition pathway from today, 2024, to 2050, and account for the heterogenous availability of renewable energy sources such as wind, solar and biomass. The analysis focuses on synergy effects and trade-offs between the industries and their respective decarbonization pathways. Our findings reveal that low-carbon hydrogen is vital in decarbonizing the chemical industry, where hydrogen is required as a feedstock in production. For high-heat processes in the steel and cement industry its application remains limited, and the utilization of biomass and the coupling with carbon capture and storage are preferred. Our integrated approach reveals that captured carbon can be utilized e.g. for methanol production, thereby reducing the pressure on limited carbon storage capacities. Results also highlight the importance of a European scale hydrogen, carbon, and biomass transport infrastructure by 2040 across all scenarios to achieve net-zero emissions by 2050.

MD77

Regency - 708

Application of Neural Networks in OR/OM

Invited Session

Computing Society

Chair: Yizhe Huang, University of Texas at Austin, Austin, TX, United States

Co-Chair: Rui Gao, University of Texas at Austin, Austin, TX, United States

1 - Deep Learning for Contextual Stochastic Optimization: Global Convergence and Statistical Guarantees

Jie Wang, Georgia Institute of Technology, Atlanta, GA, United States

Contextual stochastic optimization (CSO), by seeking the optimal decision with side information, has achieved phenomenal success in operations research. In this paper, we investigate theoretical properties of using over-parameterized two-layer neural networks for CSO. From the computational perspective, we leverage mean-field theory to show the global convergence of the training dynamics. From the statistical perspective, we provide finite-sample guarantees of this framework, and showcase its ability to overcome the curse of dimensionality and generalization to unseen data. Our theoretical results are exemplified in the Sepsis prediction task as well as strategic pricing.

2 - Policy Gradient Method for Risk-Sensitive Distributional Reinforcement Learning with Provable Convergence

Minheng Xiao, Ohio State University, Columbus, OH, United States, Xian Yu, Lei Ying

Risk-sensitive reinforcement learning (RL) is crucial for maintaining reliable performance in many high-stakes applications. While most RL methods aim to learn a point estimate of the random cumulative reward, distributional RL (DRL) seeks to estimate the entire distribution of it \citep{bellemare2017distributional}. The distribution provides all necessary information about the reward and leads to a unified framework for handling various risk measures in a risk-sensitive setting. However, developing policy gradient methods for risk-sensitive DRL is inherently more complex as it pertains to finding the gradient of a probability measure. This paper introduces a policy gradient method for risk-sensitive DRL with general coherent risk measures, where we provide an analytical form of the probability measure's gradient. We further prove the local convergence of the proposed algorithm under mild smoothness assumptions. For practical use, we also design an approximate distributional policy gradient algorithm based on categorical DRL policy evaluation \citep{bellemare2017distributional} and trajectory-based gradient estimation. Through experiments on a stochastic cliff-walking environment, we illustrate the benefits of considering a risk-sensitive setting in DRL.

3 - Unlocking the Power of Deep Q-Learning in Operations Management

Yao Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Shao-Bo Lin, Tao Li, Shaojie Tang, Ding-Xuan Zhou

With the help of massive data and rich computational resources, deep Q-learning has received unprecedented success in numerous operations management applications including recommender systems, supply chains management among others. However, there lacks solid theoretical foundations to demonstrate the success. The aim of this work is to present a unified theoretical analysis framework for Q-learning, including oracle inequalities to show the role of hypothesis space, expressive power analysis of deep neural networks in capturing special features of optimal Q-functions and generalization error analysis to quantitatively describe the relationship between size of data and prediction accuracy. Our results reveal that the effectiveness of deep Q-learning lies in the remarkable ability of deep neural networks in capturing some special properties of rewards, namely, spatial sparseness and piecewise constancy. Our theoretical assertions are verified by applying deep Q-learning in three typical applications including the well-known beer game in supply chain management, a simulated recommender system, and the Gridworld game.

4 - Learning Optimal Inventory Policies by Sequential Neural Networks

Zhen Yang, University of Texas at Austin, Austin, TX, United States, Yizhe Huang, Rui Gao

Inventory management problem is crucial for many business and have been extensively studied by previous works. Despite these effort, there remain models that are hard to solve in the sense that the structure of the optimal inventory policy remains unknown. We propose to approximate the policy at each stage by a 2-layer neural network which is a universal approximator, and show convergence results.

MD78

Regency - 709

Machine Learning Meets Discrete Optimization

Invited Session

Computing Society

Chair: Thiago Serra, Bucknell University, Lewisburg, PA, United States

1 - Learning Decision Trees and Forests with Algorithmic Recourse

Ken Kobayashi, Tokyo Institute of Technology, Meguro-ku, Japan, Kentaro Kanamori, Takuya Takagi, Yuichi Ike

In this talk, we propose a new algorithm for learning accurate tree-based models while ensuring the existence of recourse actions. Algorithmic Recourse (AR) aims to provide a recourse action for altering the undesired prediction result given by a model. Typical AR methods provide a reasonable action by solving an optimization task of minimizing the required effort among executable actions. In practice, however, such actions do not always exist for models optimized only for predictive performance. To alleviate this issue, we formulate the task of learning an accurate classification tree under the constraint of ensuring the existence of reasonable actions for as many instances as possible. Then, we propose an efficient top-down greedy algorithm by leveraging the adversarial training techniques. We also show that our proposed algorithm can be applied to the random forest, which is known as a popular framework for learning tree ensembles. Experimental results demonstrated that our method successfully provided reasonable actions to more instances than the baselines without significantly degrading accuracy and computational efficiency.

2 - Overcoming the Optimizer's Curse: Obtaining Realistic Prescriptions from ReLU Neural Networks

Asterios Tsiourvas, MIT, Cambridge, MA, United States, Georgia Perakis

We study the problem of obtaining optimal and realistic prescriptions when using ReLU networks for data-driven decision-making. In this setting, the network is used to predict a quantity of interest and then is optimized to retrieve the decisions that maximize the quantity (e.g. find the best prices that maximize revenue). However, optimizing over-parameterized models often produces unrealistic prescriptions, far from the data manifold. This phenomenon is known as the Optimizer's Curse. To tackle this problem, we model the requirement for the resulting decisions to align with the data manifold as a tractable optimization constraint. This is achieved by reformulating the highly nonlinear Local Outlier Factor (LOF) metric as a single linear or quadratic constraint. To solve the problem efficiently for large networks, we propose an adaptive sampling algorithm that reduces the initial hard-to-solve optimization problem into a small number of significantly easier-to-solve problems by restricting the decision space to realistic polytopes, i.e. polytopes of the decision space that contain at least one realistic data point. Experiments on publicly available networks demonstrate the efficacy and scalability of our approach.

3 - Optimization Over Trained Neural Networks: Taking a Relaxing Walk

Thiago Serra, Bucknell University, Lewisburg, PA, United States

Besides training, mathematical optimization is also used in deep learning to model and solve formulations over trained neural networks for purposes such as verification, compression, and optimization with learned constraints. However, solving these formulations soon becomes difficult as the network size grows due to the weak linear relaxation and dense constraint matrix. We have seen improvements in recent years with cutting plane algorithms, reformulations, and an heuristic based on Mixed-Integer Linear Programming (MILP). In this work, we propose a more scalable heuristic based on exploring global and local linear relaxations of the neural network model. Our heuristic is competitive with a state-of-the-art MILP solver and the prior heuristic while producing better solutions with increases in input, depth, and number of neurons.

Monday, October 21, 4:00 PM - 5:15 PM

ME01

Summit - 320

AI and ML for Marketplaces

Invited Session

Service Science

Chair: Zhiwei Qin, Eva AI (foreva.ai), San Jose, United States

1 - Enhancing ETA Reliability at Lyft

Rachita Naik, Lyft, Inc., New York, NY, United States

In the dynamic and competitive landscape of ride-hailing services, the accuracy of Estimated Time of Arrival (ETA) forecasts displayed prior to ride requests is crucial for ensuring customer satisfaction and fostering platform loyalty. Lyft addresses the challenge of predicting accurate ETAs prior to request time by leveraging Machine Learning (ML) to enhance ETA precision. Our strategy employs a classification approach using a tree-based ML Model that integrates a diverse dataset encompassing historical ride data, real-time traffic conditions, driver availability, and marketplace dynamics to enable nuanced predictions of ETA reliability. This methodology facilitates the selection of ETAs that align with strict reliability service levels, thereby improving the likelihood of riders choosing Lyft for their transportation needs. This session will explore the challenges of predicting ETA reliability, detail the architecture and training process of our model, and our ongoing effort to maintain and improve our predictive models and strategies in a dynamic environment, in order to uphold our commitment towards meeting user expectations and delivering service information that is both trustworthy and precise.

2 - SERGI: Similar Entity Retrieval Using Grouped Images

Akshit Sarpal, Walmart, San Bruno, CA, United States

Image data is frequently organized in semantic groups for entities such as e-commerce products, social media users or hotels on travel websites. There are a wide range of applications for retrieving entities based on their images, yet its exploration remains limited. Signals from images are commonly infused with other attributes in form of embeddings, but purely leveraging groups of images for retrieval is relatively unexplored. Drawing inspiration from natural language literature, we developed an efficient and scalable method, SERGI (Similar Entity Retrieval using Grouped Images), for retrieving entities similar to given image groups. We apply SERGI to an e-commerce use-case, aiming to identify products with brand misrepresentation. Despite the scarcity of benchmark methods for comparison, our system demonstrates superior performance compared to a baseline and a commonly used representation-based method, showing high precision in this relatively uncharted domain.

3 - A Deep Learning Approach for Predicting B2B Customers' Online Behavior

Tolunay Alankaya, Eindhoven University of Technology, Eindhoven, Netherlands, Nevin Mutlu, Sarah Gelper, Fred Langerak

As digital marketplaces become central to B2B transactions, the vast amount of data generated on e-commerce platforms offers pivotal insights into customer retention and churn. This study leverages customer clickstream data from a B2B digital marketplace to predict visits and purchases on the e-commerce platform while specifically focusing on the early detection of customer churn. Our model sequentially connects customers' clicks from previous visits to uniquely identify the customers at risk of disengaging from the platform. We use a deep learning modeling approach that improves retention and churn prediction and decreases the number of required previous visits to make an accurate prediction compared to statistical benchmarks. Moreover, the model enables us to test the effect of certain click paths on online customer behavior. Our findings contribute to academic and practical fields by emphasizing the crucial role of clickstream data in predicting B2B customer decisions. Insights from our study help managers anticipate what to expect during the digital transition and underscore the importance of data collection on B2B platform adoption. Thus, our research not only advances scholarly knowledge but also serves as a practical guide for B2B companies moving toward digital commerce.

ME02

Summit - 321

Starting Academic Career for New and Prospective QSR Faculty

Panel Session

Quality, Statistics and Reliability

Co-Chair: Xiaoyu Chen, University at Buffalo, Buffalo, NY, 14051, United States

Co-Chair: Wenmeng Tian, Mississippi State University, Mississippi State, MS, United States

1 - Moderator Panelist

2 - Panelist

3 - Panelist

4 - Moderator Panelist

Xiaoyu Chen, University at Buffalo, Buffalo, NY, United States

ME03

Summit - 322

Advancement in High-dimensional Data Analytics

Invited Session

Quality, Statistics and Reliability

Chair: Shenghao Xia, University of Arizona, Tucson, AZ, United States

Co-Chair: Jian Liu, University of Arizona, Tucson, AZ, United States

1 - Semi-Supervised PARAFAC2 Decomposition for Medical Decision Making

Elif Konyar, University of Florida, Gainesville, FL, United States, Mostafa Reisi

High-dimensional data is common in healthcare applications, particularly from the collection of data during clinical visits of patients. This HD data, commonly referred to as electronic health records (EHR), consists of medical conditions and clinical progress of a large cohort of patients over time. Tensor factorization methods are commonly used to extract meaningful patterns from the EHR data. However, challenges arise in analyzing EHR data because of data irregularity due to varying hospital visits and frequencies, and the presence of missing values, such as incomplete test results or erroneous entries. Existing PARAFAC2 frameworks are unsupervised methods and cannot exploit partially labeled data. In this work, we propose a semi-supervised PARAFAC2 decomposition model to extract meaningful patterns from irregular EHR data by also incorporating label information from a subset of instances. Experiments on synthetic and real-world data sets show the superiority of the proposed approach over the benchmarks.

2 - Federated Learning on Distributed and Encrypted Data for Smart Manufacturing

Timothy Kuo, The Pennsylvania State University, state college, PA, United States, Hui Yang

Industry 4.0 drives exponential growth in the amount of operational data collected in factories. These data are commonly distributed and stored in different business units or cooperative companies. Such data-rich environments increase the likelihood of cyber attacks, privacy breaches, and security violations. Also, this poses significant challenges on analytical computing on sensitive data that are distributed among different business units. To fill this gap, this paper presents a novel privacy-preserving framework to enable federated learning on siloed and encrypted data for smart manufacturing. Specifically, we leverage fully homomorphic encryption (FHE) to allow for computation on ciphertexts and generate encrypted results which, when decrypted, match the results of mathematical operations performed on the plaintexts. Multi-layer encryption and privacy protection reduce the likelihood of data breaches while maintaining the prediction performance of analytical models. Experimental results in real-world case studies show that the proposed framework yields superior performance to reduce the risk of cyber-attacks and harness siloed data for smart manufacturing.

3 - High Dimensional Packages' Delivery Times Prediction under Aggregated Quantile Constraints

Yifei Yuan, Amazon.com, Santa Clara, CA, United States, Mederic Motte, Philip Kaminsky

One of the fundamental aspects of the on-line shopping experience is the delivery promise – the estimated order delivery date that is displayed to customers at checkout before they purchase a product. Predicting fast and accurate packages' delivery times helps improving customers' satisfaction and attracting more purchases. For a better customer experience, the predicted delivery date should maintain a high On-time delivery rate (OTDR), which is defined as the percentage of units delivered before the predicted/promised delivery dates. OTDR metric is meaningless for a single package, because one package either meet the predicted/promised delivery date with OTDR 100% or miss the predicted/promised delivery date with OTDR 0%. Instead, there might be different OTDR requirements at different aggregated granularities, e.g. ship-option level, region level, etc. To predict delivery times under different aggregated OTDR constraints, we propose a two-step model, which first predicts delivery times' distribution at package-level and then optimize the predictions with aggregated quantile constraints. The model is tested with historical data, showing a significant reduction of the gap between promised/predicted delivery date and the actual delivery date, while maintaining the OTDR constraints.

4 - Fatigue Reliability of Manufacturing Tools Using Sensor-Based Image Sequences and Material Microstructure Analysis

Pengyan Lu, Wayne State University, Detroit, MI, United States, Qingyu Yang

The fatigue life of manufacturing tools faces extreme challenges as the strength of materials continuously increases, especially for high-strength steel. In this research, a new experimental method was designed to simulate the degradation of manufacturing tools based on image sequences. We proposed a novel methodology to study the fatigue life of both manufacturing tools and repaired tools, considering both image sequences and material microstructure information. The developed methodology was applied and verified for forming dies used with advanced high-strength steel in the auto industry.

ME04

Summit - 323

Towards More Environmentally Friendly Operations

Contributed Session

Chair: Fangling Zhu, Old Dominion University, Norfolk, VA, United States

1 - The Relationship Between It Spending in Industry 5.0 and Vertical/Horizontal Integration in Us Manufacturing

Fangling Zhu, Old Dominion University, Norfolk, VA, United States, Ling Li

This paper explores the relationship among the spending of Information technology (IT) of industry 5.0, vertical and horizontal integration of the supply chain, and firms' performance. The moderates are the competition and demand uncertainty. The explanation of the findings are follows: (1) IT spending, especially artificial intelligence, blockchain, robots, Internet of Things (IoT), digital twins, is significantly and positively associated with the vertical integration supply chain of manufacturing firms; (2) IT outsourcing and operation costs per full-time employee are positively associated factors that shows in the firms with vertical integration supply chain; (3) the effects of the vertical integration supply chain is positively associated with firms' performance, however, there is no empirical support for horizontal integration supply chain and firms' performance; (4) when uncertainty demand increases, the vertical integration of supply chain is positively correlated with firms' performance, especially for return on assets (ROA).

2 - Enhancing Sustainability in Manufacturing with Machine Learning Ensembles: A Global Empirical Study

Narges Mashhadi Nejad, University of Toledo, Toledo, OH, United States, Paul Hong, Iman Khazrak

In response to the escalating environmental and social challenges faced by the manufacturing sector, there is a growing need for robust methodologies that integrate sustainability into core operations. This paper investigates the efficacy of ensemble machine learning models, such as Random Forest, Boosting, and XGBoost, in optimizing Sustainability Value Articulation (SVA) and enhancing the execution of sustainability business models (SBMs). By analyzing data from a comprehensive survey involving 692 manufacturing firms across 23 countries, this study explores how these advanced predictive models improve upon traditional statistical methods by uncovering complex patterns, optimizing processes, and offering predictive insights into sustainability practices. The research highlights how ensemble methods facilitate a more detailed understanding of the dynamics between various sustainability practices and their impacts, enabling firms to make informed strategic decisions. Additionally, the study assesses the interactions between these predictive models and traditional factors like

Supplier Involvement Practices (SIP) and Technology Systems Integration (TSI), evaluating their combined impact on social, environmental, and economic performance indicators. The expected outcomes include a theoretically and empirically substantiated framework that demonstrates the enhanced accuracy and adaptability of ensemble machine learning in refining sustainability outcomes. This framework provides actionable insights that help businesses not only meet comprehensive sustainability goals but also effectively engage stakeholders in their sustainable operations.

3 - A Note on Drivers and Obstacles to Technology Diffusion and Market Penetration

Makiko Nagaya, Showa Women's University, Tokyo, Japan, Akira Maeda

This study examines drivers and obstacles to the diffusion of new technology. In the economics and management literature, the most commonly cited obstacles to widespread use of a new technology are cost issues, including investment costs, market imperfections and barriers to entry as well as social and psychological factors. This addresses that if the new technology is cost-effective more than before, it will naturally be accepted and implemented to the market. The realty that we observe is, however, not always the case. Cost should be understood as just one factor and there must be more fundamental mechanism.

We focus on underlying obstacles to technology diffusion beyond typical cost discussion. As a model, we consider a simple partial equilibrium of supply and demand. Using basic tools such as consumer surplus, producer surplus, and social welfare, we examine the consequences of use of a new technology through comparative statics analysis. The main results are as follows: Whether or not a new technology is adopted and diffused is determined not by technological factors on the producer side, but only by price elasticity on the demand side. If price elasticity of demand is less than 1, producers have no motivation to promote the new technology while, for any price elasticity, consumers and society as a whole will always benefit from the new technology. The result helps explain why an innovative technology sometimes face difficulties in its market penetration.

4 - Sustainable Circular Supply Chain Network in the Garment Industry: a Bi-Objective Optimization Approach with Supplier Diversity Considerations

Mayssam Tarighi Shaayesteh, University of North Texas, Denton, TX, United States

Justice and social responsibility are the two main pillars in establishing an equitable and sustainable society. Supplier diversity is a social issue that can have a substantial impact on expanding justice and promoting social responsibility. Recently, many businesses have put cooperation with diverse suppliers on their agenda to attain sustainable development goals. In this study, for the first time, a new bi-objective mixed-integer linear programming (MILP) model is formulated to configure a sustainable circular supply chain network by considering the supplier diversity problem. The first objective function (OF) of the proposed model considers the economic facet of sustainability and focuses on minimizing the total costs. The environmental facet of sustainability is included through the minimization of CO2 emissions in the second OF. In addition, cooperation with diverse suppliers is considered as the social facet. We utilize the augmented epsilon constraint method to solve the presented bi-objective MILP model, and validate its efficacy by applying the data of a garment industry in Iran. The results achieved through the practical application of the model affirm its robustness and viability as a framework in configuring a sustainable circular supply chain network.

ME05

Summit - 324

Locational Analysis for Supply Chain Design

Contributed Session

Chair: Bahram Alidaee, University of Mississippi, School of Business, University, MS, 38655, United States

1 - Solution for The Inbound Load Building-Transporting-Scheduling at Walmart

Ti Zhang, Walmart, Sunnyvale, CA, United States, Ming Ni, Liqing Zhang, Mingang Fu, Nadere Mansouri, Li Ji, Rohan Prakash, Yu Zheng, Pritam Aher, Phani Nandiraju, Bharat Dhanekula, Simon Guo, Ankush Bhargava, Jyothi Madallapalli, Amritayan Nayak, Sunil Singh, Juan Gomez, Rajiv Bhuta, Parvez Musani, Srinivasan Venkatesan

At Walmart inbound transportation, everyday tens of thousands of shipments are moved from multiple thousands of geographically dispersed vendors into Walmart's supply chain network nodes. The goal is to develop an optimization engine that reduces transportation cost, increases truck utilization, and improves on-time service qualities. The mathematical problem is a combination of a variation of set partitioning and a vehicle routing problem with time windows. Walmart's network large scale network makes this problem especially challenging to solve. We deployed a solution by transforming the problems into a divide-and-conquer framework, which employs multiphase solution strategies and algorithms: (1) network partition, a data-driven spatial partitioning scheme to divide network into non-independent subnetworks; (2) load generation, a column generation based solution framework gradually determines shipment profitability and utilizes various tailored metaheuristics including Simulated Annealing and Tabu Search algorithms to create promising candidate loads; (3) load selection, employs a combination of solution methods to select a set of loads from a large pool of candidates; (4) carrier allocation, which optimizes the carrier to load assignment while balancing carrier procurement compliance, carrier service level and transportation cost. Since the pilot launch, this optimization engine has achieved sizable business impacts: (1) better shipment consolidation that reduces trailer usage and reduces miles traveled; (2) improvement in carrier assignment to drive higher level of service; (3) transparency for root cause analysis/resolutions that improves user experience.

2 - Enhancing Walmart's Digital Fulfillment: A Multi-Channel Last Mile Network Design Framework

Amin Aghalari, Walmart, Dallas, TX, United States, Yao Luo, Xinyue Peng, Lingxun Kong, Weifang Li, Sai Mahabhashyam

Optimizing Walmart's last mile delivery (LMD) presents a complex challenge crucial for meeting its diverse, multi-channel demand efficiently. This paper introduces a comprehensive multi-channel network design framework, innovating beyond traditional single-channel planning methods. Our approach integrates scheduled and unscheduled orders, leveraging a variety of last mile nodes including store-attached automated pickup and delivery (APD), delivery stations (DS), dark stores, and direct deliveries from fulfillment centers (FC) and sortation centers (SC) to achieve significant cost savings, enhanced resource utilization, and operational efficiency.

Central to our framework is a two-phase design process. The first phase employs a mixed-integer programming (MIP) model to optimize the node and catchment decisions, minimizing total costs including last-mile delivery, fulfillment, capital costs, and penalties for unmet demand, while balancing node capacity. An innovative hierarchical pixel framework facilitates accurate last mile delivery cost estimation, avoiding the complexities of traditional vehicle-routing problems. The second phase addresses operational optimization in a dynamic environment, introducing mode-mix and last mile delivery simulation models to refine vehicle allocation and routing, ensuring network resilience under varying demand and supply conditions. This holistic framework not only streamlines Walmart's LMD operations but also underpins strategic asset investment and operational decisions, marking a leap towards enhancing customer service and operational efficiency.

3 - Decentralized Online Order Fulfillment in Omni-Channel Retailers

Sinem Savaser, Rotman School of Management, Toronto, ON, Canada, Opher Baron, Andre Cire

We consider an order fulfillment problem of an omni-channel retailer that ships online orders from its distribution center (DC) and brick-and-mortar stores. Stores use their local information to accept or reject fulfillment requests of online orders. We investigate the problem of sequencing requests to stores and inventory rationing decisions at the DC to minimize expected costs under uncertain store acceptance behavior. First, assuming stores are used only when the DC has insufficient inventory, we propose a Markov Decision Process formulation. We show that the performance rate of a myopic approach that orders stores by cost only depends on the number of items in an order, which is small in practice. Using optimality conditions for a special case of the problem, we develop an adaptive degree-based strategy that balances shipping costs and acceptance probabilities. Numerical testing suggests that the best-performing sequencing policy is within 1% of optimality on average. Moreover, using two years of data from a large omni-channel retailer in North America, we observe that adaptive policies, albeit more complex, are beneficial in reducing costs and split deliveries if acceptance rates can be estimated accurately. Second, we determine when the retailer should ship from stores or ration the inventory at the DC. We show that for single-unit orders, the optimal policy has a threshold structure, where the highest priority region is also subject to rationing. We then leverage the structure of this model to develop a heuristic for the multi-unit rationing problem and numerically establish its efficacy.

4 - Multi-Echelon Facility Location Supply Chain Design

Bahram Alidaee, University of Mississippi, University, MS, United States, Haibo Wang

At k-level uncapacitated facility location, ensuring each client is served by a sequence of k facilities is common in complex supply chain design systems. Due to its complexity, when $k > 1$, approximation algorithms are crucial for swiftly obtaining high-quality solutions. This study focuses on a k-echelon supply chain design facility location problem with $k = 5$. The objective is to select markets, distribution centers, warehouses, plants, and suppliers to maximize the total profit of serving selected markets. A Tabu Search with an r-Opt sequence embedded multi-start, and random-key diversification is designed for solution. Extensive experimental results is presented.

ME06

Summit - 325

Empirical and Experimental Analysis of Online Retailers

Contributed Session

Chair: Amin Aslani

1 - Customer Payment Choices in Virtual Reality Retail: Implications for New Sensory Environment in the Metaverse

Yoonsun Jeong, University of Texas at El Paso, El Paso, TX, United States, Nandini Nim, Jessica Felix, Leah Smith

This research explores the nature of customers' payment and transaction method choices in virtual reality (VR) retail settings. Payments present a customer touchpoint that can create a customer experience problem without proper ubiquity and interoperability across various VR and non-VR retail channels. Blockchain-based payments have gained momentum as customer payment methods in the metaverse, leading banks, financial institutions, and technology firms to develop capabilities to provide retailers and customers with more efficient and secure payment technologies. In the past decade, digital wallets such as Apple Pay have become popular across multiple retail channels and have become a preferred method. We study customer payment preferences among – credit cards, digital wallets, platform currency, and cryptocurrency. Using a VR grocery store and online grocery shopping simulation, we study the impact of an immersive sensory environment. Over two experiments and a qualitative study, we find that customers are more likely to choose digital wallets (Study 1) and platform currency (Study 2) over cryptocurrency across both VR and non-VR retail settings. We build a case for retailers and other stakeholders to build blockchain-based digital wallet capabilities with existing technology firms and financial institutions to provide customers with familiarity in the transactions and payments phase of the customer journey in VR retail journeys.

2 - How do gender and ethnic diversity impact consumer returns?

YU Jiang, University of Tennessee, Knoxville, Knoxville, TN, United States, Michael Galbreth, Paolo Letizia, Paolo Roma

In this work we study the impact of gender and ethnicity on consumer returns of online purchases. Our investigation is based on analysis of a unique dataset from a leading global luxury brand operating in the U.S. market. Our analysis confirms that gender and ethnicity both have a significant effect on consumer returns.

3 - Integrated showroom locating, pricing, and inventory management: models and analyses

Dincer Konur, Texas State University, San Marcos, TX, United States

To overcome customers' hesitation to buy online without trying a product, many online-only brands expand with showrooms. A showroom is a physical zero-inventory outlet where customers can try a product; and, if it is a fit, purchase is completed online. Showrooms are shown to increase demand. In this study, we analyze an online-only brand's showroom locating problem. Considering the demand enhancement effect of locating showrooms, pricing and inventory control decisions are integrated with showroom locating decisions. First, we formulate the integrated showroom locating, pricing, and inventory control problem as a mixed-integer nonlinear programming model. We then analyze showroom location decisions given price and pricing decision given showroom locations. These are then used to develop a polynomial time algorithm to solve the problem. The model is then extended to account for shipping fees, borne by either the customers or the brands. We show that both cases can be solved in polynomial time. Furthermore, capacitated versions of the models with demand capacity or budget limit are analyzed. For the capacitated models, we develop linear-relaxation based ranking heuristic approaches. Numerical tests indicate that the heuristic methods are computationally efficient and they return near-optimal solutions.

4 - Leveraging Augmented Reality and Assortment Planning in Omnichannel Retail Systems

Amin Aslani, University of Wisconsin - Whitewater, Whitewater, WI, United States, Borzou Rostami, Mojtaba Araghi

One of the challenges that online retailers face is the lack of physical interaction in online shopping, which can deter customer purchases. To overcome this, several major retailers are adopting augmented reality (AR) technology to enrich online shopping experiences by blending digital imagery with real-world views. This innovation enhances the customer experience through more interactive and accurate purchase decisions. Additionally, in an omnichannel retail framework, physical stores serve dual roles as both sales and information channels, where product assortments enable direct customer interaction with showcased products, complementing online AR enhancements and potentially influencing sales. Our research focuses on identifying the optimal investment in AR technology for online platforms and determining the best product assortment for physical stores to maximize overall retail system profitability. We develop solution algorithms and provide managerial insights into how these decisions are interconnected.

5 - Insights into Returned Orders: the Power of Explainable Analytics

Arno De Caigny, IESEG School of Management, Lille, France

Understanding the dynamics of orders returned by customers is crucial for firms given the significant economic losses and ecological impacts associated with these returns. Traditionally, the focus has been on managing the handling of returns rather than on proactive strategies to minimize them. Previous research has primarily concentrated on predicting the total volume of returns, which aids in managing product returns but does not give insights for preventive actions. This study shifts the focus from reactive to proactive by providing specific insights that could help firms reduce the risk of returns. Leveraging explainable analytics and a comprehensive real-world dataset from a major European e-tailer, we show how specific order- and customer- characteristics contribute to the risk of returns. Our framework offers e-tailers a strategic tool to support decision-making systems aimed at mitigating return risks.

ME07

Summit - 327

Autonomous Driving and Transportation Systems

Flash Session

Chair: Zahra Ashrafi, York University, 240 Bantry Ave, Richmond Hill, ON, L4B 4H9, Canada

1 - Adaptive Decision-Making in Human-AI Teaming for Dynamic Driving Environments

Mayuree Binjolkar, University of Washington, Seattle, WA, United States, Yuanjie Tu

This research focuses on the application of adaptive decision-making techniques, specifically bandit algorithms, to facilitate human-AI teaming in dynamic driving environments. As autonomous vehicles become increasingly prevalent, ensuring safe and efficient interactions between human drivers and AI systems remains a critical challenge. Bandit algorithms offer a promising approach to address this challenge by enabling intelligent systems to adapt their decision-making strategies in real-time to changing environmental conditions. We explore various aspects of bandit algorithms tailored for non-stationary environments, including sliding window approaches, online change-point detection, and contextual bandits. By incorporating contextual information such as traffic conditions, road dynamics, and the behavior of surrounding vehicles, these algorithms can enhance decision-making in human-AI teaming scenarios. Moreover, by actively learning from interactions with the environment, bandit algorithms contribute to the development of adaptive and intelligent systems for automated driving. In terms of specific formulation, the bandit algorithms are designed to adjust the exploration-exploitation trade-off dynamically based on an evolving reward structure, which reflects real-time data from the vehicle’s sensors and external sources. The algorithms apply a weighted combination of recent outcomes within a sliding window to prioritize current trends over older data, and use contextual cues to adapt strategies more effectively to specific driving scenarios. Through this research, we anticipate improvements in navigation precision and decision-making efficiency, particularly in complex traffic conditions. These outcomes will demonstrate the practical benefits of adaptive decision-making in improving real-world driving performance and ensuring safer human-AI interactions on the road.

2 - Leveraging Large Language Models for Autonomous Vehicle Operations

Yuming Chang, Georgia Institute of Technology, Atlanta, GA, United States, Chaojie Wang, Qing Tang, Yongyang Liu, Yangjiao Chen, Srinivas Peeta

Autonomous vehicles (AVs) seek to enhance their responsiveness to diverse driving scenarios, especially corner cases, to ensure safety and emulate human driving behaviors. However, unlike human drivers who can react to unusual traffic situations by leveraging experience and inherent skills in such situations, AVs rely on learning models which need extensive real-world data for training. The low frequency and long-tail distributions associated with corner cases makes it impractical and time-consuming to collect sufficient data through real-world field tests and simulations, posing challenges for improving AV operational performance. Recently, Large Language Models (LLMs) have illustrated the potential to conduct human-like thinking and perform tasks beyond their training data. By leveraging LLMs, we develop a generalizable methodology for creating an autonomous driving system to generate maneuvers for diverse traffic scenarios and emulate human behaviors and assess its performance using real-world corner cases. The LLM-based autonomous driving system consists of three core modules: (i) the LLM-based perception module that complements the traditional perception module by generating descriptions of corner cases from the driver's perspective, (ii) the LLM-based planning module, which generates a list of potential maneuvers by integrating the scenario descriptions with information from other sensors, and simulates trajectories to verify the maneuvers' safety, and (iii) the control module, which executes maneuvers confirmed as safe. The LLM-based autonomous driving system is tested using multiple real-world corner cases to validate its usability and reliability, in terms of how AVs respond to unfamiliar driving scenarios.

3 - Simulation Based Optimization of Autonomous Vehicle Model Parameters in Mixed Traffic by Genetic Algorithm

Jinkun Lee, East Carolina University, Greenville, NC, United States

As we observe self-driving technology of SAE level 3 and above are being tested in multiple public-private partnerships, it is generally anticipated that the autonomous vehicles (AVs) will penetrate current human driving transportation network quickly in the near future. As a result, a mixed traffic with human driving vehicles and AVs will show transient system behavior with respect to the penetration rate of AVs. Therefore we may require adaptive optimal AV settings for the overall traffic performance. We are interested in understanding this system behavior over the transitional period to achieve an optimal traffic performance. We investigate the system behavior with an agent-based simulation with different penetration rates and find optimal AV model parameter set per penetration rate by using genetic algorithm (GA). Simulation results with optimal AV parameter values reveal improvement in multiple average traffic performance measures.

4 - Navigating in Mixed-Flow Traffic: Role of Connected and Autonomous Vehicles in Shaping Human Lane-Change Behavior

Yongyang Liu, Georgia Institute of Technology, Atlanta, GA, United States, Srinivas Peeta

Connected and autonomous vehicles (CAVs) and human-driven vehicles (HDVs) will coexist on roads in the future, creating a mixed-flow traffic environment. The heterogeneity in HDVs' lane-change behaviors can significantly impact CAV operations and degrade CAV control performance. This driving simulator-based study investigates HDV lane-change behavior in mixed-flow traffic and analyzes differences in their lane-change behaviors in interactions with HDVs and CAVs. Behavior evolution is also factored as HDVs' lane-change behaviors may evolve after human drivers gain more experience in driving alongside CAVs. Results from driving simulator experiments involving 72 participants indicate significant differences in participants' lane-change behaviors in interactions with HDVs and CAVs. In addition, the learning effects from participants' interactions with CAVs on their lane-change behavior were found be significant. The study provides valuable insights on the evolution of human drivers' lane-change behaviors and the dynamics of CAV-HDV interactions in mixed-flow traffic, illustrating practical implications for developing human-friendly CAV operational systems.

5 - Vehicular Control for Improving Safety and Sustainability in Mixed Traffic of Connected Automated and Human-driven Vehicles

Brian Park, University of Virginia, Charlottesville, VA, United States

Modern vehicles are equipped with adaptive cruise control (ACC) that allows longitudinal control of the ego vehicle based on the speed or distance measured to its preceding vehicle. Some limitations include (i) the ego vehicle takes off when the preceding vehicle changes its lane, and (ii) the ACC requires 2+ seconds of time headway to be string stable, which often induces cut-ins by nearby lane vehicles. Due to these, many drivers do not use the ACC in their day-to-day driving.

When human drivers follow their preceding vehicle, their acceleration/deceleration are not optimal, mainly due to human perception and reaction. They typically do too much or too little acceleration/deceleration. Given more than 50% of new vehicles have connectivity (e.g., GM's OnStar), the ego vehicle will likely run into a connected preceding vehicle. Thus, there is a need to design and develop an algorithm that can use the connectivity from its preceding vehicle and help a human driver in charge of their driving improve their driving. To this end, a human-in-the-loop connected cruise control (hCCC) algorithm that takes advantage of vehicle connectivity and keeps drivers in control of their vehicles was developed. The hCCC takes the driver's acceleration/deceleration, receives the intended acceleration/deceleration of its connected preceding vehicle, and then adjusts the ego vehicle's acceleration/deceleration to be string stable. The hCCC was implemented and evaluated using a virtual driving simulator using an open-source Carla platform. A preliminary result showed significant reductions in fuel consumption and safety distance over human driving alone.

6 - Optimal nonlinear distance toll with mixed equilibrium of human-driven vehicles and connected and autonomous vehicles considering equity issue

Haoming Yang, Carnegie Mellon University, Pittsburgh, PA, United States

The rise of connected autonomous vehicles (CAVs) offers a promising solution to bridge the gap between user equilibrium (UE) and system optimal (SO), enhancing network efficiency and addressing the selfish routing behavior of human-driven vehicles (HDVs). However, before the period of fully CAVs appears, there will be a situation in which both HDVs and CAVs are present. In a mixed autonomy environment, we can control CAVs to follow the SO routing principle to minimize total system travel costs, while HDV users dynamically adjust their routes through day-to-day dynamics to minimize their individual travel costs in response to the CAV routing strategy. This study first solves a mixed equilibrium of HDVs and CAVs, where the mixed flow traffic assignment problem is formulated as a mathematical program with equilibrium constraints (MPEC). One challenge is that the SO routing strategy may result in higher travel times for CAVs compared to HDVs. To address this inherent unfairness, road pricing is used to influence the route choices of HDVs. We designed a piecewise nonlinear distance-based pricing scheme to further improve equity, measured with the Gini coefficient. We formulated the pricing strategy as a bi-level optimization problem, solved using artificial bee colony algorithms. The results demonstrate that the designed pricing scheme, combined with CAV routing in mixed traffic flow, can reduce inequity and total travel time in the network, even when the market penetration rate of CAVs is low.

7 - Impacts of On-board Automated Vehicle Sensors on Traffic Flow: Analysis of Sensor Range, Rate, and Resolution

Zahra Ashrafi, York University, Richmond Hill, ON, Canada, Mehdi Nourinejad, Sina Bahrami

Connected automated vehicles (CAVs) can process and interpret their environment, allowing them to navigate, identify obstructions, and make real-time decisions based on data collected by on-board sensors. CAV sensors are characterized by three features: range, rate, and resolution, commonly referred to as the *3Rs*. Range refers to the sensor's distance perception, rate to the frequency of data capture, and resolution to the data precision. The type and properties of the on-board sensors can affect traffic flow as the CAVs rely on the retrieved sensor data for making driving decisions. This study investigates the impact of the three sensor Rs on traffic flow in congested and uncongested traffic conditions. The proposed model employs car-following principles to express the relationships within the fundamental diagrams of traffic flow and their corresponding properties. The analyses show enhanced range, rate, and resolution increase traffic capacity. However, we observe a nuanced relationship between reaction time and capacity, with a unimodal pattern emerging. While an initial increase in reaction time provides CAVs with additional sensor data, thereby boosting capacity, prolonged reaction times may eventually lead to decreased capacity due to excessive caution.

This research illuminates the critical role of sensor technology in shaping traffic dynamics and highlights the potential benefits of optimizing sensor capabilities for future transportation systems.

ME08

Summit - 328

2024 MAS Annual Awards

Invited Session

Military and Security

Chair: Brian Lunday, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH, 45433, United States

Co-Chair: Shaun Doheney, Amazon Web Services (AWS), Arlington, VA, United States

1 - Mas President's Introductory Remarks

Shaun Doheney, Amazon Web Services (AWS), Herndon, VA, United States

This session is made possible by the efforts of MAS Past-president, Dr. Brian Lunday, who orchestrated the MAS Annual Awards for 2024, along with the chairs of the respective committees: Dr. David Myers for the Bonder Scholarship; Dr. Tahir Ekin for the Koopman Prize; Dr. Ed Pohl for the J. Steinhardt Prize; and Dr. Matthew Robbins for the MAS Student Paper Competition.

2 - Mas Student Paper Competition

Matthew Robbins, Air Force Institute of Technology, WPAFB, OH, United States

See https://connect.informs.org/militaryandsecurity/awards/studentpapercompetition

3 - Seth Bonder Scholarship for Applied Operations Research in Military and Security Applications (Sponsored by the Bonder Estate) - Presentation by Recipient on Emerging Research Thread

Shaun Doheney, Amazon Web Services (AWS), Herndon, VA, United States, David Myers

See https://connect.informs.org/militaryandsecurity/awards/seth-bonder-scholarship

4 - Koopman Prize (Sponsored by MAS Members)

Ekin Tahir, Texas State University, San Marcos, TX, United States

See https://connect.informs.org/militaryandsecurity/awards/koopman-prize

5 - J. Steinhardt Prize (Sponsored by Cna Corporation) - Presentation by Recipient on Career Insights and Observations

Ed Pohl, University of Arkansas, Fayetteville, AR, United States

The J. Steinhardt Prize is sponsored by the CNA Corporation. The prize is awarded for outstanding contributions to Military Operations Research and is awarded for life work rather than for any particular contribution. The selection Committee is composed of previous award winners. The award is accompanied by a plaque and a $2,000 honorarium.

ME09

Summit - 329

Matching Markets

Invited Session

Auctions and Market Design

Chair: Sophie Yu, Stanford University, Durham, NC, United States

Co-Chair: Pengyu Qian, Boston University, Boston, United States

1 - Incentives Design for Decentralized Matching

Pengyu Qian, Boston University, Boston, MA, United States, Chen Chen

In decentralized dynamic matching markets, agents can benefit from sharing their supply and demand, but they will collaborate only if properly incentivized. Motivating applications include kidney exchange and ridesharing platforms. This paper explores incentive design to foster collaboration in such markets. We consider a model with multiple self-interested agents, each operating a general multi-way dynamic matching market, aiming to maximize their time-average matching rewards. The agents' states are private information, and items awaiting matches are available for only a limited amount of time. We introduce a simple reward splitting mechanism that specifies the distribution of rewards for inter-player matches, leading to a complex dynamic game. Despite its complexity, we demonstrate that in all market equilibria, the total reward closely approaches that of a centralized matching market, with the gap diminishing to zero as the number of players increases to infinity. This mechanism is straightforward to implement and offers significant economic insights.

2 - Experiment Design for One-Sided Matching Platforms

chenran weng, University of Science and Technology of China, Hefei, China, People's Republic of, Xiao Lei, NIAN SI

One-sided matching markets, where users are matched with other users, are evident in environments like video game platforms and anonymous social networks. Here, participants are matched for interactions such as games or social exchanges. Experimentation (A/B tests) in these markets is challenging due to the interdependence of users' metrics on their counterparts' treatment assignments. In this paper, we build a stochastic market model and develop its mean field limit to analyze such experimental dynamics. Our focus is on two randomization strategies: user and match randomization. We demonstrate that, under Markovian conditions and homogeneous users behavior, match randomization provides unbiased estimations but can lead to significant biases when these conditions are not met. Conversely, user randomization shows greater resilience to model inaccuracies. We further propose an associated linear regression estimator that can halve the bias compared to a naive estimator.

3 - Stochastic Online Metric Matching: Adversarial is No Harder than Stochastic

Mingwei Yang, Stanford University, Stanford, CA, United States, Amin Saberi, Sophie Yu

We study the stochastic online metric matching problem. In this problem, n servers and n requests are located in a known metric, where all servers are available upfront and requests arrive one at a time. In particular, servers are adversarially chosen, and requests are independently drawn from a known distribution. Upon the arrival of a new request, it needs to be immediately and irrevocably matched to a free server, resulting in a cost of their distance. The objective is to minimize the total matching cost. The best competitive ratio achieved for general metrics is $O((\log \log \log n)^2)$, and only the tree metrics are known to admit $O(1)$-competitive algorithms.

In this paper, we show that the problem can be reduced to a more accessible setting where both servers and requests are drawn from a fixed distribution by losing only a constant factor in the competitive ratio. Combined with previous results, we give $O(1)$-competitive algorithms for a broader range of metrics and distributions. For $[0, 1]^d$ with $d \geq 3$, we present an $O(1)$-competitive algorithm for distributions satisfying a mild condition. We further extend our algorithm to the case with at most a linear amount of excess servers and achieve the same competitive ratio guarantee.

4 - Stable Matching as Transport

Joseph Root, University of Chicago, Chicago, IL, United States, Federico Echenique, Fedor Sandomirskiy

We study matching markets with aligned preferences and establish a connection between common design objectives—stability, efficiency, and fairness and the theory of optimal transport. Optimal transport gives new insights into the structural properties of matchings obtained from pursuing these objectives, and into the trade-offs between different objectives. Matching markets with aligned preferences provide a tractable stylized model capturing supply-demand imbalances in a range of settings such as partnership formation, school choice, organ donor exchange, and markets with transferable utility where bargaining over transfers happens after a match is formed.

ME10

Summit - 330

Selected Auctions & Market Design papers from the EC 2024 conference I

Award Session

Auctions and Market Design

Chair: Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Ali Makhdoumi, Duke University, Durham, NC, United States

Co-Chair: Pengyu Qian, Boston University, Boston, United States

1 - Pure Nash Equilibrium Market Recommendations

Shai Vardi, Purdue University, West Lafayette, IN, United States

New business models help sellers make better decisions by communicating information about market prices. However, sellers' inability to coordinate greatly reduces the efficacy of this information and can lead to market failures. We study the feasibility and benefits of providing equilibria as recommendations in economies where sellers face price uncertainty, information scarcity, and an inability to coordinate. We describe a general model in which sellers wish to sell some quantity of a good at one of several markets with elastic prices, and show that while an equilibrium may not necessarily exist, an approximate equilibrium always exists under mild assumptions on the market concentration. %provide an algorithmic framework that a market planner can use to recommend ex-post equilibria to sellers. We describe an algorithm that a market planner can use to recommend approximate ex-post equilibria to sellers, and compare the recommended (approximate) equilibria to selling strategies identified from a survey of onion and potato farmers in India. Using both synthetic data and real data on India's agricultural markets, we show that the recommended strategy outperforms the other strategies in terms of seller welfare, geographic price dispersion, market volume concentration, and several metrics of individual farmer welfare improvements. We also show that the framework is robust to imperfect market penetration, is fair to smaller farmers, and captures nearly all of the welfare from a welfare-maximizing allocation.

2 - Contract Design With Safety Inspections

Alireza Fallah, UC Berkeley, Berkeley, CA, United States

We study the role of regulatory inspections in a contract design problem in which a principal interacts separately with multiple agents. Each agent's hidden action includes a dimension that determines whether they undertake an extra costly step to adhere to safety protocols. The principal's objective is to use payments combined with a limited budget for random inspections to incentivize agents towards safety-compliant actions that maximize the principal's utility. We first focus on the single-agent setting with linear contracts and present an efficient algorithm that characterizes the optimal linear contract, which includes both payment and random inspection. We further investigate how the optimal contract changes as the inspection cost or the cost of adhering to safety protocols vary. Notably, we demonstrate that the agent's compensation increases if either of these costs escalates. However, while the probability of inspection decreases with rising inspection costs, it demonstrates nonmonotonic behavior as a function of the safety action costs. Lastly, we explore the multi-agent setting, where the principal's challenge is to determine the best distribution of inspection budgets among all agents. We propose an efficient approach based on dynamic programming to find an approximately optimal allocation of inspection budget across contracts.

3 - Strategically-Robust Learning Algorithms for Bidding in First-Price Auctions

Rachitesh Kumar, Amazon, Seattle, WA, United States

Learning to bid in repeated first-price auctions is a fundamental problem at the interface of game theory and machine learning, which has seen a recent surge in interest due to the transition of display advertising to first-price auctions. Past work has predominantly treated it as an online learning problem and focused on developing regret-minimizing algorithms. However, such an approach ignores the strategic nature of the market in which such algorithms are deployed---sellers and buyers can manipulate these algorithms for their own benefit. In this work, we go beyond regret minimization and develop bidding algorithms that are robust to manipulations by strategic market participants. We propose a novel concave formulation for pure-strategy bidding in first-price auctions and use it to analyze Gradient Ascent. Unlike previously-proposed algorithms, which attain optimal regret but are manipulable, we show that Gradient Ascent attains optimal regret while being robust to strategic manipulations by the buyer and the seller.

4 - Information Design and Pricing of Supply Competition in Lead-Selling Platforms

Yanwei Sun, Imperial College Business School, London, United Kingdom

We analyze a type of platform common in the service sector, where customers can freely post tasks, and service providers pay a feeto purchase job leads and unlock the ability to communicate with these customers. We adopt a Bayesian persuasion framework to explore how the platform can leverage its informational advantage about the supplier base and pricing to strategically influence the decisions of service providers, aiming to maximize its revenue. We begin the analysis by assuming that prices are exogenously given and characterize the structure of the optimal signaling mechanism. Specifically, the platform recommends all service providers to participate when the pool-size is below a certain threshold. When the pool-size exceeds this threshold, the platform randomizes its recommendations between two consecutive numbers, independent of the pool-size. This behaviour is driven bya unique feature of our model, where the beliefs of the platform and suppliers are non-identical: a supplier's knowledge of their own participation inflates their belief about the total number of service providers. Next, we endogenize the platform's pricing decision and consider the policy space where the signaling mechanism and state-dependent pricing are jointly optimized. Notably, within this space, we findthat the optimal signaling mechanism with a single price remains optimal. This result implies the superiority of the optimal signaling mechanism over the commonly adopted contingent pricing policy, where the price charged by the platform depends on the number of interested suppliers. We show that the performance discrepancy between these two policies is particularly pronounced when supply competition is intense.

ME11

Summit - 331

Emerging Trends in Financial Risk Management and Market Influence

Contributed Session

Chair: Yusaku Watanabe, N/A

1 - Necessity Analysis of Insurance and Reinsurance in Mitigating Bank Loan Risk -- Based on Multi-Party Game Model

Jinye Du, Southern University of Science and Technology, Shenzhen, China, People's Republic of, Yanhong Guo

The financial vulnerabilities of Small and Medium-sized Enterprises (SMEs) is a vital problem and insurance and reinsurance can be an efficient way of risk mitigation. By employing a game theory framework, our study examines interactions among SMEs, banks, insurers, and reinsurers, focusing on strategic interdependencies and the complexity of financial support mechanisms. Distinct from previous research centred on profitability, our analysis highlights the challenge of transparency in borrowers' financial integrity during initial loan approvals, a factor clarified in subsequent oversight. The proposed model addresses SME financing challenges and efficiently distributes risks among all stakeholders—including enterprises, financial institutions, reinsurance companies, and government regulators—thereby fostering a sustainable financial ecosystem. Our approach enhances banks' lending capacity while effectively managing default risks, contributing to the stabilisation of the national economic order.

2 - Impact of Firm-Level Cyber Risk on Stock Price Informativeness: An Empirical Analysis

Waqas Nawaz Khan, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Shan Liu, Jae Kyu Lee

Cyber risk is a critical aspect of the economy all over the globe. The relationship between firm-level cyber risk and stock price informativeness has become a critical area of research, particularly in an era where cyber threats loom large over corporate governance. This study investigates how the exposure to cyber risk at the firm level influences the informativeness of stock prices in the market. Utilizing a robust dataset of 26252 firms from 2007-2018 that spans several industries, we employ multiple regression analyses to examine the impacts of cyber risk on the efficiency of stock price in reflecting relevant information. Our findings indicate that higher levels of cyber risk are significantly associated with stock price informativeness. This suggests that as firms become more susceptible to cyber threats, the efficient stock prices. Stocks of firms with higher cyber security risk tend to achieve higher informativeness both from managers' and fundamental investors' efforts. High cyber security risk leads to lower earnings management and increased fundamental investors' information acquisition. Our path analyses suggest that decreased earnings management and increased fundamental investors' attention are the key mechanisms of impact through which high stock price informativeness is achieved. The research contributes to the ongoing discussions on corporate transparency and market uncertainty surrounding their future cash flows and operational stability increases, leading to less efficiency, highlighting the importance of robust cyber risk management strategies not only for safeguarding operational integrity but also for maintaining investor confidence and market stability.

3 - Reference Price Effect Model as the Piecewise-Smooth Dynamical System: Dynamic Pricing with Asymmetric Reference Price Effects

Anton Bondarev, Xi'an Jiaotong-Liverpool University, Suzhou, China, People's Republic of, Regis Chenavaz

In this paper we contribute to the study of the well-known reference price effects model, established by Fibich et al. (2003) and subsequent papers.

We make use of recently developed methods in hybrid control of piecewise-smooth dynamical systems to shed light to several aspects of that model which previously were not studied.

We start with the classical symmetric price effects model and next analyze asymmetric effects both under loss aversion and gain seeking assumptions. In the first case we are able to fully replicate previous findings and extend them claiming that there is an additional continuum of equilibria in between two (previously reported) price levels associated with different reference price effects.

For the gain seeking case, we extend the literature by showing that there are additional optimal solutions associated with the existence of the sliding equilibrium and its associated flow. This result is to our best knowledge, novel to the field.

At last we extend our findings to the differential game setup with many players and show that main results survive such an extension. We also discuss further potential development sin the field.

4 - Estimating the Cost of Content Diversification: An Empirical Evidence from Japanese TV Markets

Yusaku Watanabe, Kobe University, Kobe, Japan

Diversity in broadcasting is an important concept that has been set as a primary goal of broadcasting policy in many countries. There are also many studies on media diversity in various academic fields. However, while diversity in media is recognized as important from a social and policy perspective, most studies have overlooked the discussion of the costs associated with content diversification. Scheduling programming is an essential strategic decision for broadcasters and needs to be treated endogenously.

The purpose of this study is to address this gap by examining the impact of broadcasters' content diversity on viewer and advertiser demand, and estimating the marginal cost of content diversification using a structural model based on a two-sided market model. This study utilizes a market-level dataset for the Japanese TV market, including advertising prices, advertising quantities, audience ratings, and detailed program characteristics such as content diversity, to estimate the proposed model. Viewer choice behavior is formulated as a Multiple Discrete-Continuous Extreme Value (MDCEV) model, in accordance with the frequency of data observations, and transformed to be estimated at the market-level. The study shows that the marginal cost can be obtained by combining the structural model and estimated demand parameters.

By quantifying the costs of achieving diversity in broadcasting content, this study offers insights that can inform more effective policymaking to promote a vibrant and pluralistic media landscape.

ME12

Summit - 332

Radiotherapy and Cardiovascular Health

Flash Session

Chair: Jindong Tong, University of Florida, Gainesville, FL, United States

1 - Determining the Causal Influence of Migraine-Specific Factors on MACE Outcome: A Causal Learning Approach

Asra Aghaei, Arizona State University, Tempe, AZ, United States, Teresa Wu, Catherine Chong

Migraine is a common and debilitating headache disorder that has a one-year prevalence of 12% in the general population. Existing data demonstrate that migraine is associated with an increased risk of cardiovascular disease and stroke, which are commonly referred to as MACE events. The NIH's *All of Us* research program includes over 647,000 participants with approximately 80% of those originating from underrepresented groups. As such, it is one of the most diverse health databases in history and includes participants from demographic groups that are typically underrepresented in human health research. With the *All of Us* database, 34,232 individuals received a migraine diagnosis (female=27,409/80%, male=5,997/18%, other=726/2%; mean age=54.1, std=15.6) and 379,225 did not have a migraine diagnosis (female=222,056/59%, male=149,172/39%, other=7,997/2%; mean age=54.9, std=17.2). We queried ICD-9 and ICD-10 codes for migraines and strokes and compared the prevalence of strokes between individuals with and without migraines. Utilizing relative risk ratios with a 95% confidence interval, results demonstrate Ischemic stroke was 3.38 times higher (95% CI: 3.25 to 3.52; $P < 0.001$), hemorrhagic stroke was 3.18 times higher (95% CI: 2.93 to 3.44; $P < 0.001$) and ill-defined stroke was 5.31 times higher (95% CI: 4.80 to 5.87; $P < 0.001$) in individuals with migraine compared to those without migraine. These results suggest migraines are associated with a higher prevalence of both ischemic and hemorrhagic stroke. This research presents our preliminary study on casual analysis to investigate the cascading relationships amongst migraine-related factors which influence MACE outcome.

2 - Survival Analysis for Risk Stratification of Clinical Events in Heart Failure Patients Using Longitudinal EHR Data

Maryam Jafaripakzad, University of South Florida, Tampa, FL, United States, Jorge A. Acuna Melo, Jose L. Zayas Castro, Joel Fernandez

Heart failure affects approximately 6.5 million Americans over the age of 20, directly causing about 8.5% of all heart disease deaths in the United States. It is also a leading cause of hospitalizations among the Medicare population, emphasizing the urgent need for improved management and predictive strategies. Accurately predicting clinical events in heart failure patients is essential for proactive healthcare management. Advanced predictive models enable clinicians to identify at-risk patients earlier and more reliably, facilitating timely interventions that can prevent deterioration of health and avoid costly hospital admissions. In this study, we aim to predict adverse clinical events in heart failure patients by using longitudinal electronic health record data, which includes repeated measurements of risk factors over time. We applied and compared dynamic prediction models, including time-varying Cox regression, joint modeling of longitudinal and time-to-event data, and recurrent neural networks (RNNs) to assess their predictive accuracy and their ability to adapt to new information. Preliminary analyses indicate that these dynamic models outperform traditional static models in predicting clinical events, particularly in survival data.

3 - Deep Learning-based Bio-signal Multivariate Time-series Forecasting

Seonyeong Park, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

Heart disease remains a significant health concern, accounting for a substantial portion of mortality rates worldwide. This underscores the critical importance of early diagnosis and prediction in improving patient outcomes and survival rates. Given the importance of bio-signal data in diagnosing heart conditions, particularly in the context of heart disease, the ability to forecast bio-signal patterns emerges as a critical problem. Recurrent Neural Networks (RNNs) such as Long-short term memory(LSTM) and Gated Recurrent Unit(GRU), designed to handle sequential data with a memory mechanism, are particularly well-suited for modeling temporal dependencies in bio-signal records. By using LSTM and GRU, our study enhances the ability to forecast multivariate time series with bio-signal patterns. We used various open real-world multivariate time series datasets provided by UCI machine learning dataset, PhysioNet and etc.. Various preprocessing methods suitable to bio-signal datasets are used, and time series preprocessing methods are used, too. Through the RNN-based model and various time series preprocessing methods, our study enables the forecasting of patients' future bio-signal patterns based on past bio-signal data. The experimental results of our method showed higher accuracy in forecasting bio-signal patterns than the other models. Our methods not only provide the effectiveness, but also insights to help in the decision-making process of heart disease prevention and treatment strategies. Analyzing bio-signal patterns by our method can provide valuable insights related to healthcare, and help to develop proactive management and prevention strategies for heart disease.

4 - Enhancing Heart Failure Diagnosis: A Transparent Approach with Ensemble of Simple Models

Gaurav Arwade, Iowa State University, AMES, IA, United States, Sigurdur Olafsson

Heart failure remains a significant global health issue, with 6.5 million confirmed cases in the United States alone and an annual increase of 960,000 new cases. The complex nature of this clinical syndrome, often intertwined with various other health conditions, presents challenges in achieving accurate diagnoses. Thus, certain ischemic and valvular heart conditions contribute to heart failure being overlooked. Conversely, symptoms that overlap with conditions such as obesity and respiratory diseases can result in both over and underdiagnoses. Breathlessness, a common symptom shared between heart failure and other conditions, can lead to misdiagnosis, particularly in primary care settings where comprehensive evaluation resources are scarce. Predictive models have the potential to enhance diagnostic accuracy, but transparency and explainability are essential in this field. To address these diagnostic challenges, we propose an ensemble approach that relies on simple models constructed based on the inherent structures of the data. These models utilize subsets of patients with similar characteristics to minimize redundant complex interactions among features. Though deep learning and complex algorithms like XGBoost show promising diagnostic capabilities, their lack of explainability raises physicians' concerns about 'black box' recommendations. algorithms. Our approach prioritizes improved explainability without compromising accuracy, making it more accessible and reliable for clinicians. We demonstrate how the simple structures within the ensemble differ from each other and the global model using causality. This further enhances our understanding of the model's predictions and aids in interpretation, strengthening its utility in clinical practice for precise heart failure prediction in primary care settings.

5 - A Mathematical Framework for Optimizing Immuno-radiotherapy Treatments

Ehsan Salari, Wichita State University, Wichita, KS, United States, Keerthi Venugopal Rakhee

Immunotherapy has recently emerged as an important modality for the treatment of advanced-stage cancers. Despite the encouraging clinical results, the number of patients who durably respond to immunotherapy alone is small. Therefore, there is clinical interest in combining immunotherapy with radiotherapy to achieve better outcomes. The primary goal of this study is to develop novel mathematical models to inform the design of immuno-radiotherapy (IRT) treatments. To that end, we adapt and utilize existing compartmental models available in the literature to explicitly optimize for the dosing and schedule of IRT treatments. Specifically, we consider an extended compartmental model based on differential equations to simulate the population dynamics of cancer and immune cells under the influence of radiation and immune-checkpoint inhibitors (ICI's). We integrate the compartmental model into an optimization framework to optimize the antitumor immune response while limiting the risk of normal-tissue toxicity. The proposed IRT optimization framework determines both a radiation fractionation schedule and an ICI dosing schedule. Additionally, it allows for investigating how variations in disease characteristics such as cancer stage, primary tumor volume, tumor radio-sensitivity, tumor growth rate, and metastatic drivers, may affect the optimal IRT schedule.

6 - Using Zeroth Order IMPT Algorithm Towards Clinic Implementation of the MKM RBE model for Carbon Ion therapy

Jindong Tong, University of Florida, Gainesville, FL, United States, Dong Han, Hongcheng Liu, Yu Yang, Bo Lu

Ion beam radiotherapy, especially using protons and carbon ions, has recently gained significant attention in the field of radiation oncology. The Microdosimetric Kinetic Model (MKM) model utilizes novel expressions to calculate exposure-specific cell survival, considering the full spectrum of microdosimetric information instead of relying on average values. While the MKM model shows promise for describing the

Relative Biological Effectiveness (RBE) in Carbon Ion Radiotherapy (CIRT), its integration into treatment planning for RBE dose optimization has been less visited due to its sophisticated computations involved in the model that prohibits easy access to improving search directions. To address these challenges, we adopt random gradient free method for tractable optimization.

ME13

Summit - 333

Supply Chain Risk Assessment

Contributed Session

Chair: Mohamad Darayi, The Pennsylvania State University--Great Valley, Malvern, PA, United States

1 - Value of Information in Analyzing Freight Transportation Resilience Considering Economic Impacts

Mohamad Darayi, The Pennsylvania State University--Great Valley, Malvern, PA, United States

Defining a four-phase disaster life cycle as (i) mitigation, (ii) preparedness, (iii) response, and (iv) recovery, freight transportation resilience literature most focus is on assessing the network's ability to deal with disruption consequences, with less work on holistic analysis of resilience. This work, however, emphasizes resilience analysis considering economic impacts. A systematic framework to analyze freight transportation resilience is proposed by (i) evaluating vulnerability of the interdependent system of freight infrastructure and economic sectors, (ii) seeking strategies to allocate a limited budget to harden the system, and (iii) assessing value of information in decision making under uncertainty. The proposed resilience analysis framework is implemented in a stylized case study of the multi-modal freight transportation network that connects Pennsylvania to the surrounding states.

2 - "Effects of Heterogenous Decisionmakers in Disrupted Supply Chains"

Souri Sasanfar, Northeastern University, Boston,, MA, United States

Title:

"Effects of Heterogenous Decisionmakers in Disrupted Supply Chains"

Disruptions to the supply chain present serious problems, especially in the pharmaceutical industry where reliable and effective delivery is essential. This study examines the effects of decision-maker heterogeneity on supply chain performance during different types of disruptions, with a particular emphasis on three important profiles: followers, hoarders, and reactors. To study the dynamic interactions between these decision-makers and two important information sharing formats, Automated System Information (ASI) and Automated Demand Information (ADI), we created an agent-based simulation model.

The simulation looks at how different decision-maker types react to supply chain interruptions and how their actions interact with ADI and ASI to impact performance. By lowering supply chain expenses, panic purchases, and phantom ordering, the study seeks to determine which information sharing practices could enhance performance and offer suggestions for the best kinds of ADI and ASI for pharmaceutical supply chains.

The goal of the research is to provide a comprehensive understanding of the actions and tactics used by various decision-makers, along with recommendations for improving collaboration and decreasing inefficiencies. By doing so, stability in this important industry will be ensured and pharmaceutical supply chains will become more robust and effective.

3 - Revisiting Supply Chain Response Strategy Under Uncertainty and Asymmetric Information

chuqiao peng, The University of Oklahoma, Norman, OK, United States, Qiong Wang

In supply chain management, uncertainty and asymmetric information pose significant challenges. However, the unique impacts of these challenges on supply chain response strategies remain unclear. Using secondary datasets, this study empirically explored how firms employ supply chain response strategies—specifically, bridging and buffering—to navigate these challenges. Limited the sample size to be publicly traded companies, we are able to control for the firm level variables using Compustat. Utilizing FactSet Revere Relationship dataset, this study measures the supply chain response strategies through the inter-organizational relationship between buyers and suppliers and uses the location as a proxy for asymmetric information. Employing WRDS SEC Readability and Sentiment Data, we are able to explore the supply chain uncertainty using sentimental analysis in the 10-K filing. Additionally, the lengthy time series of these datasets enables us to mitigate the impacts of external policy changes and unobservable noises on the estimation. Drawing upon Information Processing Theory, this study conceptually reveals that uncertainty intensifies the implementation of bridging and buffering strategies. In contrast, asymmetric information generally undermines these strategies' effectiveness. These insights offer valuable managerial implications, underscoring strategic approaches for navigating the complexities of modern, often volatile, business environments. While uncertainty and asymmetric information are often viewed as volatile and challenging, they can be transformed into valuable opportunities for growth and success when firms adopt the appropriate response strategies.

4 - Non-committed cost auditing in supply chain contracts

Jing Li, Renmin University of China, Beijing, China, People's Republic of, Wanshan Zhu, Zhengping Wu

As a widely used means in supply chain practice, audit can be used as an auxiliary means to alleviate the supply chain incoordination problem under asymmetric information that cannot be solved by profit-sharing contract and other means. Although some studies have considered audit issues under supply chain contracts in the literature, few studies are conducted from the perspective of principal-agent theory, not to mention the situation where retailers do not commit to audit strategies under the principal-agent model. In this paper, under the principal-agent theory model, we consider the optimal audit strategy when the retailer makes no commitment to the audit mechanism. We explore the unique features of the audit strategy and the sensitivity of the optimal strategy to various factors.

ME14

Summit - 334

Quantitative Finance and FinTech

Invited Session

Finance

Co-Chair: Chen Yang, The Chinese University of Hong Kong, Sha Tin, Hong Kong

1 - Implementing a Ramsey Plan

Wei Jiang, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

We study how to induce future governments to continue a fiscal plan chosen by a benevolent government at time 0, an issue analyzed by Lucas and Stokey (1983). We implement a Ramsey plan by adding instantaneous debt to the contractible subspace and requiring each continuation government to preserve that debt's purchasing power over the next instant. After first using a Lagrangian to solve the Ramsey problem, we formulate the problem recursively and then apply that formulation to settings with various debt term structures and government spending processes. We extract implications about tax smoothing and e effcts of fiscal policies on bond markets.

2 - Risk-Sharing Pricing of Variable Annuities within a Principal-Agent Framework

Bin Li, University of Waterloo, Waterloo, ON, Canada, Min Dai, Dongchen Li, Yumin Wang

We propose a new risk-sharing pricing approach of variable annuities within a principal-agent framework where an insurer (principal) is the contract provider and a policyholder (agent) is the follower having the surrender option. While the risk-neutral pricing approach adopted in the existing literature leads to significantly higher fees and more frequent surrendering than market observations, this new risk-sharing pricing approach reconciles the misalignment between theoretical results and market observations. Our risk-sharing framework is flexible to consider the policyholder's outside options such as bond or mutual fund investments, as well as the insurer's financial friction costs. Additionally, we examine the impacts of policyholder discount uncertainty, tax-deferred benefits, and competition among multiple insurers as extensions.

3 - Reinforcement learning to optimal stopping

Yu Sun, The Hong Kong Polytechnic University, Kowloon, Hong Kong, Min Dai, Zuoquan Xu, Xunyu Zhou

We study the optimal stopping problems with unknown model primitives in the continuous-time reinforcement learning framework developed by Wang et al. [Journal of Machine Learning Research 21(1):8145–8178, 2020]. By penalty approximation of the HJB equation, we transform the optimal stopping problems to optimal control ones, for which we randomize the control at each time by a probability measure and introduce an entropy regularizer to encourage exploration. We derive the semi-analytical optimal Bernoulli distribution of the decision regarding stopping or not, and devise reinforcement learning algorithms based on the martingale approach developed by Jia and Zhou [Journal of Machine Learning Research 23(154):1–55, 2022]. Finally, we demonstrate the algorithms in an example of finite-horizon American put option, and both the offline and online algorithms achieve high accuracy in terms of pricing the option and characterizing the associated free boundary.

4 - Optimal Tax-Timing with Transaction Costs

Chen Yang, The Chinese University of Hong Kong, Shatin, Hong Kong, Min Dai, Yaoting Lei, Hong Liu

We develop a dynamic portfolio model incorporating capital gains tax (CGT), financial transaction tax, and transaction costs, where the tax amount is calculated at the end of each year. We find that transaction costs affect loss deferrals much more than gain deferrals, and a lower interest rate makes higher-wealth investors realize losses sooner but makes lower-wealth ones realize losses later. Our model can help explain the puzzle that even when investors face equal long-term/short-term CGT rates or almost zero interest rates, they may still defer realizing large capital losses. In addition, it provides several unique, empirically testable predictions and can shed light on recently proposed tax policy changes.

5 - Corporate Investment and Savings Demand

Ling Qin, Chinese University of Hong Kong, Hong Kong, Hong Kong, Nan Chen, Xavier Giroud, Neng Wang

Why do some firms save more than others? We revisit this important, widely studied question by developing a tractable continuous-time capital-accumulation model for financially constrained firms facing costly external equity. In addition to including the standard building blocks for a q theory of investment (Hayashi, 1982; Abel and Eberly, 1994): capital accumulation, capital adjustment costs, and a persistent stochastic productivity process, we assume that instantaneous profits (cashflows generated by productive capital) are stochastic. This assumption is a key difference between our model and widely used quantitative corporate finance models, e.g., Hennessy and Whited (2007) and Riddick and Whited (2009), in which instantaneous profits (conditional on productivity and capital stock) are deterministic.

Firm value depends on capital stock, cash balance, and productivity. Using the homogeneity property, we analytically characterize the solution with a variational inequality for cash-capital ratio and productivity. Importantly, we show that making instantaneous profits stochastic is key to generating quantitatively meaningful cash holdings. Consider the alternative that instantaneous profits are (locally) deterministic. Then the firm's cash balance at the end of the period would also be deterministic. This substantially reduces its precautionary savings demand, as locally the risk-neutral firm (as long as it is not too cash strapped) faces little liquidation risk and hence holding cash has limited value. Mathematically, whether (conditional) instantaneous profits are stochastic implies whether the second-order derivative of firm value with respect to cash appears in the Bellman equation, and economically this term has a first-order effect on precautionary savings demand.

ME15

Summit - 335

Topics in Decision-Making under Uncertainty

Invited Session

Revenue Management and Pricing

Chair: Billy Jin, Cornell ORIE, Chicago, IL, United States

1 - Online Linear Programming: the Regret Landscape of the Ce Heuristic

Yilun Chen, CUHK Shenzhen, Shenzhen, China, People's Republic of, Wenjia Wang

The Certainty Equivalent (CE) control is a canonical heuristic approach to dynamic resource allocation problems, central to the literature of OR and MS. The CE approach dynamically (re)solves for the optimal solution of a corresponding fluid problem, derived from replacing all the uncertainty by their average values, to guide its decision-making. While CE and its variants are widely applied in practice, the existing theoretical performance guarantees of CE typically require prohibitively strong condition, known as the non-degeneracy conditions. In this work, we analyze the performance of CE within the general framework of online linear programming. Under mild assumptions on problem primitives, we show that CE achieves optimal regret (up to polylog factors) even when the problem is degenerate, thus expanding the existing understanding of the effectiveness of CE for general dynamic resource allocation problems. Furthermore, we characterize the regret landscape achieved by CE in terms of a coefficient describing the smoothness of the problem's underlying probability distribution. Our results essentially relax the non-degeneracy conditions to only requiring the fluid problem to have unique solutions, which has been shown previously (in the special case of multi-secretary problem) to be necessary for CE-type heuristics to succeed.

2 - A Constant Approximation Framework for the Online-Retail Item Placement Problem

Duygu Soylemez, The University of Chicago Booth School of Business, Chicago, IL, United States, Levi DeValve

For online retailers, warehouse assortment planning significantly impacts the profitability of fulfilling multi-item orders. Once items are placed in warehouses, considering item budgets, the challenge is to determine which warehouses will satisfy each order under operational capacity constraints. We propose a modeling framework for this problem, as well as some other related problems, where we establish a novel form of approximate submodularity. This enables us to analyze local search heuristic algorithms that terminate in polynomial time and to propose a general scheme that guarantees new constant factor approximations.

3 - Understanding Group Structure in individual Decision-making

Jessica Dai, U.C. Berkeley, Berkeley, CA, United States, Paula Gradu, Nika Haghtalab, Deborah Inioluwa Raji, Benjamin Recht, Eric Zhao

When consequential decisions are made about individuals, questions of fairness or discrimination in these contexts depend critically on understanding how individuals relate to those groups. In this talk, we share two perspectives on making (online) decisions when per-group performance is fundamental to the problem. In the first, in a sequential hypothesis-testing task, we assume that the group membership of individuals are known; in this case, we show that per-group validity can be guaranteed even when groups are overlapping, and/or the number of groups is very large (possibly exponential in the number of features). In the second, in a prediction task, we show how approximate per-group calibration can be achieved even when group definitions are unknown a priori.

4 - Online Job Scheduling with Uniform Random Arrivals

Jody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States, Benjamin Moseley

Worst-case analysis of the classic single-machine online job scheduling problem with the objective of maximizing weighted throughput has seen limitations due to it being impossible to construct a constant competitive algorithm for it. All subsequent work on this problem either makes assumptions about the input including the release times, processing times, and processing window or leverages resource augmentation. In this paper, we take a beyond worst-case approach and propose the random-order temporal model that samples the release times of jobs from an uniform distribution. Using this new model, we design constant competitive algorithms for both unweighted and weighted throughput maximization.

ME16

Summit - 336

Deep Learning Application in RM

Invited Session

Revenue Management and Pricing

Chair: Srikanth Jagabathula, NYU Stern School of Business, New York, NY, United States

Co-Chair: Sandeep Chitla, NYU Stern School of Business, New York, NY, 10012, United States

1 - Consumer Purchase Behavior in Online Grocery: a Structural Modeling Approach Using Generative AI

Sandeep Chitla, NYU Stern School of Business, New York, NY, United States, Srikanth Jagabathula, Vidyadhar Mudium, Shubha Shedthikere

Understanding how customers build their shopping carts can help retailers influence what customers buy and how much they spend. At what cart value should we offer free delivery? Should we offer a cart-level discount or offer a discount on a specific product to increase total spend? To answer these questions, the retailer must understand how customers build their carts and how various factors affect the next item they add to the cart. Building on the recent success of transformer-based models, we propose a generative model to predict the next item a customer will add based on the current items in the cart and the cart value. We train our model on a large dataset obtained from our industry partner, Swiggy Instamart, which is a leading online grocery service in India. Our results showcase the accuracy of our model and its efficacy in helping the firm optimize promotions.

2 - A Robust Approach to Classification and Truth-Validation: Incorporating Human and Large Language Model Decision Making

Chaithanya Bandi, Northwestern University, Evanston, IL, United States

The advent of generative AI and foundational models has opened up promising avenues to enhance the efficiency of many service systems. However, their wider application has been stymied by the persistent presence of model errors and inaccuracies. These errors represent an intrinsic characteristic of current AI models, and not just a transient hurdle. In this talk, we present a Robust Optimization based approach that embraces this reality and seeks to optimize performance within this context.

We present a Robust optimization-based algorithmic framework and develop the Fast Algorithm for Classification and Truth-validation (FACT), which aims to address the critical need for rigorous truth-validation in service systems. FACT is designed to leverage the decision-making competencies of both humans and foundational Large Language Models (LLMs) through a dynamic task routing strategy, informed by task complexity, projected cost, and the capacity of the decision-making entity.

We present our results from a pilot implementation at a customer service portal of an online education company, which demonstrates significant efficiency benefits. This approach offers a promising trajectory for future research and practical implementations, especially in the context of systems where AI and human decision-making must be effectively integrated.

3 - Gpt2 for Operational Decision Making

Shang Liu, Imperial College Business School, London, United Kingdom, Hanzhao Wang, Kalyan Talluri, Guanting Chen, Xiaocheng Li

In this paper, we consider the supervised pretrained transformer for a class of sequential decision-making tasks. The class of considered tasks is a subset of the general formulation of reinforcement learning in that there is no transition matrix, and the class covers bandits, dynamic pricing, and newsvendor problems as special cases. Such a structure enables the use of optimal actions/decisions in the pretraining, and the usage also provides new perspectives for understanding the training and generalization of the pretrained transformer. We first note that the training of the transformer model can be viewed as a performative prediction problem, and the existing methods and theories largely ignore or cannot resolve the arisen out-of-distribution issue. We propose a natural solution that includes the transformer-generated action sequences in the training procedure, and it enjoys better properties both numerically and theoretically. The availability of the optimal actions in the considered tasks also allows us to analyze the properties of the pretrained transformer as an algorithm and explains why it may lack exploration and how this can be automatically resolved. Numerically, we categorize the advantages of the pretrained transformer over the structured algorithms such as UCB and Thompson sampling into three cases: (i) it better utilizes the prior knowledge in the pretraining data; (ii) it can elegantly handle the misspecification issue suffered by the structured algorithms; (iii) for short time horizon such as $T\le50$, it behaves more greedy and enjoys much better regret than the structured algorithms which are designed for asymptotic optimality.

4 - The Choice-Learn Python Package: a Deep Learning Lens on Choice Modeling

Vincent Auriau, CentraleSupélec, Gif-sur-Yvette, France, Ali Aouad, Antoine Desir, Emmanuel Malherbe

Discrete choice models aim at predicting choices made by individuals from a set of options, called an assortment. Such models are used as inputs to operational problems such as assortment optimization and pricing. We develop Choice-Learn, a Python package for choice modeling practitioners and researchers. The package enables processing choice data, and then formulating, estimating, and operationalizing choice models based on the TensorFlow library. We provide a unified implementation of standard parametric families of choice models as well as neural network-based methods. The code combines high-level usage, designed for fast and easy implementation, and low-level usage for customization. Basic post-processing tools that leverage choice models for assortment optimization and pricing are also integrated into the package.

ME17

Summit - 337

Online Resource Allocation: New Models and Methods

Invited Session

Revenue Management and Pricing

Chair: Rajan Udwani, UC Berkeley, Berkeley, CA, United States

1 - Random-Order Contention Resolution via Continuous Induction: Tightness for Bipartite Matching Under Vertex Arrivals

Calum MacRury, Columbia University, New York, NY, United States, Will Ma

We introduce a new approach for designing Random-order Contention Resolution Schemes (RCRS's) via exact solution in continuous time.
Given a function $c(y):[0,1] \ra [0,1]$,
we show how to select each element which arrives at time $y \in [0,1]$ with probability \textit{exactly} $c(y)$.
We provide a rigorous algorithmic framework for achieving this, which discretizes the time interval and also needs to sample its past execution to ensure these exact selection probabilities.
We showcase

our framework in the context of online contention resolution schemes for matching with random-order vertex arrivals. For bipartite graphs with two-sided arrivals, we design a $(1+\sfe^{-2})/2 \approx 0.567$-selectable RCRS, which we also show to be \textit{tight}.
Next, we show that the presence of short odd-length cycles is the only barrier to attaining a (tight) $(1+\sfe^{-2})/2$-selectable RCRS on general graphs. By generalizing our bipartite RCRS, we design an RCRS for graphs with odd-length girth g which is $(1+\sfe^{-2})/2$-selectable as $g \rightarrow \infty$. This convergence happens very rapidly: for triangle-free
graphs (i.e., $g \ge 5$), we attain a $121/240 + 7/16 \sfe^2 \approx 0.563$-selectable RCRS.
Finally, for general graphs we improve on the $8/15 \approx 0.533$-selectable RCRS of (Fu et al., 2021) and design an RCRS which is at least 0.535-selectable.
Due to the reduction of (Ezra et al., 2020), our bounds yield a 0.535-competitive (respectively, $(1+\sfe^{-2})/2$-competitive) algorithm for prophet secretary matching
on general (respectively, bipartite) graphs under vertex arrivals.

2 - The Online Submodular Assignment Problem

Daniel Hathcock, Carnegie Mellon University, Pittsburgh, PA, United States, Billy Jin, Kalen Patton, Sherry Sarkar, Michael Zlatin

Online resource allocation is a rich and varied field. One of the most well-known problems in this area is online bipartite matching, introduced in 1990 by Karp, Vazirani, and Vazirani. Since then, many variants have been studied, including AdWords, the generalized assignment problem (GAP), and online submodular welfare maximization.

In this work, we introduce a generalization of GAP which we call the submodular assignment problem (SAP). This generalization captures many online assignment problems, including all classical online bipartite matching problems as well as broader online combinatorial optimization problems such as online arboricity, flow scheduling, and laminar restricted allocations. We present a fractional algorithm for online SAP that is (1 - 1/e)-competitive.

Additionally, we study several integral special cases of the problem. In particular, we provide a (1 - 1/e - \epsilon)-competitive integral algorithm under a small-bids assumption, and a (1 - 1/e)-competitive integral algorithm for online submodular welfare maximization where the utility functions are given by rank functions of matroids.

The key new ingredient for our results is the construction and structural analysis of a “water level” vector for polymatroids. This construction reveals connections to submodular utility allocation markets and principal partition sequences of matroids.

3 - Adaptive Policies and Approximation Schemes for Dynamic Matching

Ali Aouad, London Business School, London, United Kingdom, Alireza Amanihamedani, Amin Saberi

Motivated by applications to ride-hailing platforms and organ donation schemes, we study a dynamic matching problem, where customers and servers arrive according to Poisson point processes. Customers must be immediately matched to an available server within the network of queues, and servers renege after a random duration. The goal is to optimize the long-term average throughput and cost of the matches.

Previous literature on dynamic matching often focuses on static policies—where matching decisions are not adapted to dynamic information on the servers’ queue lengths. In this work, we formulate tighter linear programming (LP) approximations that represent an adaptive matching process, where the queue lengths inform the matching decisions in real time. This new LP allows us to develop a bi-criteria fully polynomial-time approximation scheme for single-queue instances. Using a lifted polyhedral relaxation, we devise a polynomial-time approximation schemes (PTAS) for instances with a constant number of server types. By combining this approach with careful instance decompositions, we devise the first PTAS for dynamic matching on fixed-dimensional Euclidean graphs.

4 - Online Submodular Welfare Maximization Meets Post-Allocation Stochasticity and Reusability

Rajan Udwani, UC Berkeley, Berkeley, CA, United States

We generalize the problem of online submodular welfare maximization to incorporate a variety of new elements arising from reusability, stochastic rewards, combinatorial actions and similar features that have received significant attention in recent years. For our general formulation, we show that a non-adaptive Greedy algorithm achieves the highest possible competitive ratio against an adaptive offline benchmark in the adversarial arrival model and also under some assumptions in the unknown IID stochastic arrival model. This generalizes several previous results and shows that, in general, adaptivity to stochastic rewards (and similar features) offers no theoretical (worst-case) benefits.

ME18

Summit - 338

Pricing Aspects of Platform Design

Invited Session

Revenue Management and Pricing

Chair: Michael Hamilton, University of Pittsburgh, Pittsburgh, PA, United States

Co-Chair: Titing Cui, University of Pittsburgh, Pittsburgh, PA, United States

1 - Distance-Based Fee Design of Rapid Delivery

Manlu Chen, Renmin University of China, Beijing, China, People's Republic of, Mingliu Chen, Ming Hu, Jianfu Wang

This paper examines the evolving landscape of the rapid delivery service industry, focusing on the strategic variations in pricing structures and service offerings as firms expand their operations from urban to suburban areas. Utilizing a stylized queueing model, we analyze the effects of a two-part delivery price—which includes a base fee and a distance-dependent fee—alongside the availability of a self-pickup option. The study evaluates the implications of these strategies for various stakeholders across various customer density environments. In urban centers, our findings indicate that vendors target nearby customers, instituting a positive per unit distance fee to deter distant orders. This approach optimizes courier capacity for servicing nearby customers. In contrast, vendors operating in suburban settings aim to attract remote, high-valuation customers by introducing a negative per unit distance fee, effectively subsidizing delivery costs—a practice less favorable to customer welfare and social welfare. We further investigate the potential profitability of self-pickup as an alternative to delivery services. In densely populated urban areas, self-pickup proves more lucrative, leading vendors to incentivize this option through discounts. Conversely, in suburban regions with lower population densities, vendors may impose additional fees on self-pickup to favor delivery services. Our analysis concludes that, despite the varying profitability across demographic regions, the provision of a self-pickup option consistently benefits vendor profits and customer welfare. This research underscores the critical role demographic characteristics play in the strategic deployment of rapid delivery and self-pickup services, with significant implications for the operational and pricing strategies of firms within the industry.

2 - Fresh Rating Systems: Structure, Incentives, and Fees

Titing Cui, University of Pittsburgh, Pittsburgh, PA, United States, Michael Hamilton

Rating systems aggregate customer reviews into a numeric measure of service quality. However, not all reviews — and consequently, not all ratings — are equally informative. The relevance of a review diminishes as it ages, making older reviews less indicative of a service provider's current quality. Older, outdated reviews are often termed \textit{stale} and pose a number of challenges for traditional rating systems. Rating systems that do not consider the age of reviews may not accurately represent the current service quality, they are also more susceptible to manipulation, and can lead to perverse incentives for service providers.

In this paper, we explore how to modify rating systems to address the issue of stale reviews. We consider two types of staling, sequential staling and temporal staling. To mitigate issues due to sequential staling, we introduce a new class of $\al$-moving average rating systems which systematically down-weight older reviews. We show that the moving average approach generates ratings in a way which optimally protects customers from changes in the underlying quality of service that may otherwise go undetected due to staling. To address issues due to temporal staling, we extend our rating system to include penalty ratings that are incurred by long periods of inactivity. In a model of online platforms that facilitate interactions between service providers and customers, we give a prior-free way of choosing the penalty term and associated parameters which approximately maximizes revenue and also discourages users and providers from disintermediating on the platform.

3 - The Power of Two in Token Systems

Süleyman Kerimov, Rice University, Houston, TX, United States, Itai Ashlagi, Omer Tamuz

In economies without monetary transfers, token systems serve as an alternative to sustain cooperation, alleviate free riding, and increase efficiency. This paper studies whether a token-based economy can be effective in marketplaces with thin exogenous supply. We consider a marketplace in which at each time period one agent requests a service, one agent provides the service, and one token (artificial currency) is used to pay for service provision. The number of tokens each agent has represents the difference between the amount of service provisions and service requests by the agent. We are interested in the behavior of this economy when very few agents are available to provide the requested service. Since balancing the number of tokens across agents is key to sustain cooperation, the agent with the minimum amount of tokens is selected to provide service among the available agents. When exactly one random agent is available to provide service, we show that the token distribution is unstable. However, already when just two random agents are available to provide service, the token distribution is stable, in the sense that agents' token balance is unlikely to deviate much from their initial endowment, and agents return to their initial endowment in finite expected time. Our results mirror the power of two choices paradigm in load balancing problems. Supported by numerical simulations using kidney exchange data, our findings suggest that token systems may generate efficient outcomes in kidney exchange marketplaces by sustaining cooperation between hospitals.

4 - Price Parity and Competition on E-Commerce Platforms

Yi Tong, University of Florida, Gainesville, FL, United States, Jingchuan Pu, Quan Zheng

To prevent marketplace leakage, e-commerce platforms usually enforce a price parity clause on third-party sellers, forbidding them from offering lower prices elsewhere online, including their own websites. Regulators have criticized this requirement as an anti-competitive practice detrimental to sellers and consumers. In this study, we examine the marketplace's incentives for implementing price parity and investigate its impacts on sellers and consumer welfare under different market conditions, providing timely insights for policymakers. Along with the price parity, the marketplace's introduction of its own brands is also widely criticized due to the intensified competition faced by third-party sellers. Against this backdrop, we first examine the impact of the marketplace's own brand and price parity on sellers and consumers. We further analyze whether and how these two retail arms interplay.

ME19

Summit - 339

Advancements in Network Management and Analysis

Contributed Session

Chair: Feiran Xu, University of Iowa, Iowa City, IA, 52242, United States

1 - Approximation Algorithms for Directed Weighted Spanners and Buy-at-bulk Network Design

Young-San Lin, Melbourne Business School, Carlton, Australia, Elena Grigorescu, Nithish Kumar

We present efficient sublinear-factor approximation (online) algorithms for directed pairwise weighted spanners and buy-at-bulk network design, where each edge is associated with a cost and a length. The goal is to find a minimum-cost subgraph that satisfies terminal distance constraints.

2 - Convolutional Similarity Network Fusion on Multiplex Graphs

Sangheum Cho, Ajou University, Suwon, Korea, Republic of, Taehwan Yun, Hyunjung Shin

Multiplex graphs denote a collection of graphs that share the same set of nodes but possess different sets of edges, depending on the data source. As data availability grows, utilizing data with diverse source information, such as multiplex graphs, becomes increasingly important. In particular, if each graph that makes up multiplex graphs has complementary information, it can contribute to better performance of machine learning models. Despite this advantage of complementary information, recent research on graph-based deep learning models has mainly focused on dealing with single graph rather than multiplex graphs. Therefore, we propose deep learning framework specialized for multiplex graphs using the Similarity Network Fusion (SNF) algorithm, which is one of the integration methods for structured data. In our proposed method, SNF is extended to the problem of multiplex graphs by exploiting the embedding process with graph convolution. As a results of experiments on toy and real-world datasets, the proposed method showed significantly better performance than comparison methods, and especially maintained high performance even in situations where there is little labeled data.

3 - Sensor Placement Strategy Under Limited Budget for Anomaly Detection in Large Scale Networks

Feiran Xu, University of Iowa, Iowa City, IA, United States, Ramin Moghaddass

Large-scale graph-structured systems are becoming increasingly prevalent, and it is typically essential important to maintain their stability and reliability. One key challenge in ensuring this reliability is the timely detection of anomalous nodes and the identification of the original sources of disturbances. Reading and analyzing the real-time sensor data from within the network can facilitate this task. However, the limited availability of sensors poses a significant constraint, as it is impractical to deploy sensors at every node. Therefore, determining the optimal placement of a limited number of sensors to maximize system observability is a fundamental yet challenging problem. This work introduces an innovative cluster-based framework designed to identify an optimal set of nodes for sensor placements for anomaly detection within a large-scale network. The proposed method is evaluated through a series of data experiments, demonstrating its effectiveness in enhancing network observability under budget constraints.

4 - Strategic Risk and Cost Reduction in Transportation Networks: a Data-Driven Optimization Approach

Zeinab Vosooghi, McMaster University, Hamilton, ON, Canada, Manish Verma

This study presents an innovative methodology for identifying critical zones within a transportation network, with a specific focus on hazardous material (hazmat) transportation. By employing a data-driven optimization-analytics approach, the aim is to minimize both the cost and risk associated with hazmat railroad transportation problems. Initially, a comprehensive nationwide risk assessment is conducted for Class 1 railroads across the United States, utilizing data mining techniques applied to the Federal Railroad Administration (FRA) dataset on hazardous material-related accidents. This analysis facilitates the development of a robust predictive model for estimating the likelihood and damages of future accidents. Subsequently, strategic areas within the network are identified by solving a bi-objective model that seeks to minimize both network cost and risk. The integration of well-established methodologies such as weighted sum and epsilon constraint enhances the reliability of the results, which are subsequently validated.

ME20

Summit - 340

Practical Applications of Decision Analysis

Invited Session

Decision Analysis Society

Chair: luis lillo, Pontificia universidad Catolica de Valparaiso, santiago, Chile

1 - Dynamic analysis for predicting the hydraulic performance of the wine bottling system based on operational and comprehensive indicators.

Luis Lillo, Pontificia universidad Catolica de Valparaiso, Valparaiso, Chile, Hanns De la Fuente, Jose Ceroni

The wine industry in Chile has significant socio-cultural and economic importance. However, it faces challenges such as climate change, declining consumer preferences compared to other products, and a constant increase in costs, especially in energy and labor. In this context, the final filtration stage before bottling is crucial due to its high cost and variability, as well as its role in ensuring microbiological stability and organoleptic attributes through particle retention that can alter the wine's properties. Particle retention reduces the available filtering area, increases the pressure differential, and gradually decreases the flow, affecting the bottling speed to levels that may render the process unfeasible. Traditionally, the tools and strategies of the industry do not allow for timely forecasting of these situations, impacting the production plan and costs.

In response to these challenges, a machine learning model has been developed and evaluated for dynamic analysis and prediction of the hydraulic performance of the filtration and bottling system, taking into account the physicochemical properties of the wine, comprehensive indicators, and the measurement of operational parameters. This model is designed to estimate the efficiency of the filter media and the flow evolution during the bottling process, enabling the estimation of bottling time and planning of cleaning processes or line stoppages due to filter media fouling. With this, the aim is to improve operational decision-making during bottling, allowing for more effective control of costs and their variability through the use of adaptable and reliable tools with potential for industry-wide application.

.

2 - Investment challenges under renewable variability and water scarcity for the operation of microgrids.

Yolanda Matamala, Universidad Católica de la Santísima Concepción, Concepción, Chile, Kevin Melendez

Investing in Power-To-X technologies could prove infeasible considering challenges such as, 1) The future of energy development could soon be limited by water scarcity. 2) The variability of renewable sources poses challenges to the electrical grid in a scenario of growing demand and climate change events. 3) Economic viability associated with the high investment costs in flexible technologies or infrastructure. This study propose a strategic plan for the feasibility of investment in Power-To-X technologies under water scarcity scenarios for microgrid communities in connected mode. The interaction of microgrids is modeled through a stochastic bi-level model that considers water availability constraints. The results show that stochastic planning allows for more fluid electricity and hydrogen production and storage strategies, ensuring consistent operations with a narrower production range. The information obtained could be used to contribute to public policies that consider financing sustainable development for microgrid communities.

3 - A MIP Formulation for the Bottling Scheduling Problem in the Wine Industry with Sequence-Dependent Setup Times

Alonso Peña Domarchi, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, Luis Lillo, Hanns de la Fuente, Abhishake Kundu

A systematic review has revealed a gap in the wine industry: the planning of the bottling process and the complexities of the filtering stage have not been addressed simultaneously. In response to this, a novel optimization model will be formulated using Mixed Integer Programming (MIP) for the last stage of the production process. The objective is to minimize production times, considering the interdependencies of operational decisions and the specific conditions of the bottling and filtering processes in the wine industry. The model can be considered a scheduling of jobs to production lines with sequence-dependent setup times. To test the applicability of the model, data from one of the largest wineries in the Chilean wine industry will be used. Computational experiments will be developed to create an algorithm that solves the additional complexity of larger instances.

4 - A Two-Stage Stochastic Optimization Framework for Efficient Food Management in Hospitals

Mariana Arriz Jorquiera, University of South Florida, Tampa, FL, United States, Jorge Acuna, Jose Zayas-Castro

Food waste in healthcare systems presents a significant environmental and nutritional challenge, contributing up to 5% of global environmental damage and impacting patient nutrition. This study introduces a two-stage stochastic optimization framework that minimizes food cost and waste in hospital settings. In the initial stage, the model estimates food requirements for the upcoming week. Subsequently, the model builds the meals for the patients considering diverse dietary needs, demographics, and three possible demand scenarios. Our approach integrates flexibility by accounting for last-minute purchases to address shortages and managing inventory surplus efficiently. The stochastic nature of the model is driven by the probabilities associated with each scenario, determined by the distribution of the patient's length of stay. This integration of stochasticity enhances the model's resilience against unforeseen fluctuations in demand, improving the efficiency and effectiveness of food procurement and distribution in hospital settings. Overall, this framework offers economic benefits for the hospital and contributes to environmental sustainability and patient well-being.

5 - Optimizing Kidney Transplant Allocation in the US: A Two-Stage Stochastic Approach

Daniela Cantarino, University of South Florida, Tampa, FL, United States, Jorge A. Acuna, Jose Zayas-Castro

Chronic kidney disease (CKD) presents a significant challenge to the United States' population, impacting longevity and quality of life. Renal transplants offer a pivotal intervention, extending life expectancy by an average of 10 years compared to dialysis. However, the persistent shortage of viable kidneys highlights the urgent need for more efficient allocation strategies. To address this, we propose a novel multi-period two-stage stochastic model aimed at improving the quantity and quality of matches by focusing on pre- and post-transplant processes. The model accounts for uncertain parameters and evaluates key considerations such as waitlist mortality, expected survival rates, travel distances, organ degradation, and biological compatibilities. By running simulations, we calculate Quality-Adjusted Life Years (QALYs) to assess the effectiveness of different allocation strategies, providing valuable insights for enhancing the efficiency and equity of the kidney transplant system.

6 - Bilevel programming methods for hyperparameter optimization in support vector machines: A classification and regression approach

Gonzalo Ríos, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, Hanns de la Fuente-Mella, Felipe Feijoo

Support vector machine (SVM) is a classical model well known for its performance as a learning method for regression and classification tasks in a wide range of problems that require machine learning tools. In this study we introduce a novel framework with a reformulation for hyperparameter selection in SVM for support vector regressor (SVR) and support vector classifier (SVC) with bi-level programming models, both in SVR and SVC the upper-level problem minimizes the error for predicted values, considering misclassification metrics for SVC and mean absolute error for SVR, while the lower-level problems set the values for parameters and hyperparameters on the entire model. The bi-level problem is converted to a mathematical program with equilibrium constraints (MPEC) model. The MPEC is solved by applying Karush-Kuhn-Tucker (KKT) conditions over the lower–level problems and the nonlinear constraints derived from complementary conditions on KKT / MPEC are linearized by using the Big-M method and special ordered set type 1 (SOS1) variables. In experiments the models and reformulations are tested against public and simulated datasets with different shapes, showing the model reformulated and linearized through SOS1 variables is faster than the one linearized with the Big-M method. Results provide evidence, not only of the efficiency shown by the model reformulated with SOS1 variables against the Big-M method but also of the differences of a bi-level approach against classical machine learning model approaches commonly used in the professional and scientific community.

ME21

Summit - 341

Energy, Electricity Markets, and Sustainability

Invited Session

Decision Analysis Society

Chair: Fariba Farajbakhsh Mamaghani, Tulane University, 6001 Mounes St, River Ridge, LA, 70123, United States

1 - Multistorage - Power vs. Energy - When To Transition Between Storage Technologies

Christian Kaps, Harvard Business School, Cambridge, MA, United States, Simone Marinesi

As the world transitions towards a decarbonized energy future, the choice of energy storage technologies plays a pivotal role in matching the variable supply of renewable generation with demand. In this paper, we model the trade-off between different storage technologies, focusing on the difference between how much energy a technology can store and how much power it can provide at any given time. This distinction is crucial as the use cases for and cost of large-scale energy storage technologies differ.
along these dimensions in drastic ways, impacting under which circumstances they can be operated profitably.

With this paper, we aim to answer three research questions: i) What factors determine the utilization of a given storage technology; b) how does the duration for which energy needs to be stored (hours, days, weeks, seasons) impact which storage technology is most useful; c) how does the existence of multiple storage technologies impact the value or utilization compared to each technology in isolation?

2 - Renewable, Flexible, and Storage Capacities: Friends or Foes?

Owen Wu, Indiana University, Bloomington, IN, United States, Xiaoshan Peng, Gil Souza

Over 99% of the new power generation capacity to be installed in the U.S. from 2023 to 2050 will be powered by wind, solar, and natural gas. Additionally, large-scale battery systems are planned to support power systems. It is paramount for policymakers and electric utilities to deepen the understanding of the operational and investment relations among renewable, flexible (natural gas-powered), and storage capacities. In this paper, we optimize both the joint operations and investment mix of these three types of resources, examining whether they act as investment substitutes or complements. Using stochastic control theory, we identify and prove the structure of the optimal storage control policy, from which we determine various pairs of charging and discharging operations. We find that whether storage complements or substitutes other resources hinges on the operational pairs involved and whether executing these pairs is constrained by charging or discharging. Through extensive numerical analysis using data from a Florida utility, government agencies, and industry reports, we demonstrate how storage operations drive the investment relations among renewable, flexible, and storage capacities. Storage and renewables substitute each other in meeting peak demand; storage complements renewables by storing surplus renewable output; renewables complement storage by compressing peak periods, facilitating peak shaving and displacement of flexible capacity. These substitution and complementary effects often coexist and the dominant effect can alternate as costs change. A thorough understanding of these relations at both operational and investment levels empowers decision-makers to optimize energy infrastructure investments and operations, thereby unlocking their full potential.

3 - Breaking the utility death spiral: A regime-switching fee-based model for utilities firms serving customers with renewable energy generation

Alexandar Angelus, Texas A&M University, College Station, TX, United States

It is well established that utilities serving customers with distributed solar power generation face deteriorating profits due to the resulting demand reduction and frequently mandated net metering compensation for those customers' excess generated power. In this research, we propose a regime-switching fee-based mechanism that protects an electric utility from the impact of distributed renewable generation in the presence of net metering while still incentivizing customers' renewable energy investments. In the context of the infinite-horizon, sequential game played between a utility and its customer, we establish that this pricing mechanism can achieve perfect coordination of the supply chain consisting of a customer and his utility, and attain Pareto improvements in value function for both parties.

4 - Energy consumption of limited response to weather

Fariba Farajbakhsh Mamaghani, Tulane University, River Ridge, LA, United States, Saed Alizamir, Shouqiang Wang

The integration of smart technologies and automation into the power grid has empowered utility firms to monitor real-time consumption and experiment with price-based incentive mechanisms to mitigate demand-supply imbalances. However, recent implementations of Demand Response (DR) programs in residential sectors have yielded mixed results, challenging assumptions about consumer responsiveness to price signals. One major obstacle is the inability of fixed prices to account for exogenous factors like outdoor temperature fluctuations. This limitation leads to erratic energy demand patterns as households struggle to adjust consumption rationally. Addressing this, we propose a novel framework rooted in the concept of "rational inattention" to analyze households' decision-making processes. By considering cognitive limitations and information processing costs, we identify regions where households are responsive or non-responsive to demand shocks, driven by comfort or cost-saving motives. Our analysis reveals nuanced relationships between price, consumption variability, and overall market reliability. Using real data from New York City apartments, we calibrate our model to offer insights into empirical demand patterns. Our findings complement existing DR program analyses, providing utility firms and policymakers with normative recommendations for designing effective pricing mechanisms and anticipating market responses.

ME22

Summit - 342

DAS Awards Session

Award Session

Decision Analysis Society

Chair: Yael Grushka-Cockayne, University of Virginia, Charlottesville VA, VA, United States

1 - Minimax-Regret Sample Selection in Randomized Experiments

Yuchen Hu, Management Science and Engineering, Stanford University, Stanford, CA, United States, Henry Zhu, Emma Brunskill, Stefan Wager

Randomized controlled trials are often run in settings with many subpopulations that may have differential benefits from the treatment being evaluated. We consider the problem of sample selection, i.e., whom to enroll in a randomized trial, such as to optimize welfare in a heterogeneous population. We formalize this problem within the minimax-regret framework, and derive optimal sample-selection schemes under a variety of conditions. Using data from a COVID-19 vaccine trial, we also highlight how different objectives and decision rules can lead to meaningfully different guidance regarding optimal sample allocation.

2 - Axiomatizing the Bayesian Paradigm in Parallel Small Worlds

Simon French, Alliance Manchester Business School, Manchester, United Kingdom

There is currently much interest in scenario-focused decision analysis (SFDA), a methodology that provides, among other things, supporting analyses in circumstances in which there are deep uncertainties about the future (that is, when experts and decision makers (DMs) cannot come to any agreement on some of the probabilities to use in a Bayesian model). This lack of agreement can mean that sensitivity and robustness analyses show that virtually any strategy may be optimal under the beliefs of one or more participants. Scenario-focused analyses fix the deep uncertainties at interesting values in different scenarios and conduct a (Bayesian) decision analysis within each. The results can be informative to the DMs, helping them understand different possible futures and their reactions to them. However, theoretical axiomatizations of subjective expected utility (SEU), the core of decision analysis, do not immediately extend to the context of SFDA. The purpose of this paper is to provide an axiomatization of SEU that supports SFDA. Scenarios have much in common with Savage’s concept of small worlds. We discuss the parallels and then explore two difficulties in extending his and other writers’ axiomatizations. The development of SEU offered here overcomes these difficulties. Throughout, attention is given to the implications of the theoretical development for the practice of decision analysis.

3 - Reflections on a Decision Analysis Career by 2024 Frank P. Ramsey Medal Winner

Samuel Bodily, University of Virginia, Charlottesville, VA, United States

The Medal honors Ramsey, who introduced decision theory, where probability, or degree of belief was made operationally meaningful as willingness to act. This year’s awardee found this broadly applicable. He will talk about his career research, manifested in scholarly articles; his work in practice, exemplified by consulting and applied articles for business and government audiences; and his teaching, exemplified by cases and technical notes for students, articles on teaching effectiveness, and building a teaching team.

ME23

Summit - 343

Pricing and Revenue Management Problems in Supply Chain Networks

Invited Session

Revenue Management and Pricing

Chair: Chun Ye, Amazon.com, Seattle, WA, United States

Co-Chair: Goutam Kumar, Amazon, Seattle, WA, United States

1 - Bandit-based assortment strategies with application to ultra-fast delivery

Siyeon Kim, Amazon, Raleigh, NC, United States, Hunyong Cho, Jonathan Jonker, Lina Al-Kanj, Boran Hu, Greg Herman, Josiah Davis, Zachary Hervieux-Moore, Eric Laber

Ultra-fast delivery is becoming an increasing important component of e-commerce. However, an effective ultra-fast delivery offering requires a finely tuned distribution system. At the heart of this system is an assortment that is as lean as possible while meeting customer purchase intent. To ensure an assortment that adapts to evolving customer preferences and that accommodates an ever-changing catalog of possible items, the assortment must be continuously refined through judicious experimentation. We propose a bandit-based exploration strategy based on customer segmentation that ensures assortments are fresh, regional, and responsive to changing demand. We illustrate the proposed methodology through experiments with Amazon's Sub-Same Day delivery service.

2 - Dynamic Pricing via Reinforcement Learning in a Spot Marketplace.

Jacob Tutmaher, Amazon, Denver, CO, United States, Huiwen Jia, Priyanka Shende, Mohsen Moarefdoost, Philip Kaminsky

We develop reinforcement-learning-based pricing algorithms for a transportation load board. The algorithm modifies the prices of available loads dynamically to achieve a set of performance goals, such as minimizing the operation costs, encouraging the use of clean energy equipment, and attracting enough carriers to secure the transportation needs. The optimal pricing policy is closely related to carriers’ willingness to pay and thus can be affected by several elements, such as the overall status of freight industry and the existence of substitutable loads or competing carriers interested in the same load. We formulate this dynamic pricing problem as a Markov Decision Process (MDP) and solve for a dynamic pricing policy via offline reinforcement learning.

3 - Offer Selections Optimizer in Transportation Capacity Sourcing

Yaser Ghaedsharaf, Amazon, Bellevue, WA, United States, Mohsen Moarefdoost, Huiwen Jia, Philip Kaminsky

A marketplace platform is used to source last-minute transportation loads. If a load is not covered on the platform, it goes to a manual sourcing channel where an agent engages in negotiations with carriers. The current negotiation process is manual and lacks scientific guidance, leading to higher costs. This paper formulates the problem as a Markov Decision Process and develops a machine learning-based model to optimize the negotiation process, resulting in significant cost savings.

4 - Spot Market Pricing on Amazon Freight

Goutam Kumar, Amazon, Seattle, WA, United States, Adam Elmachtoub, Roger Lederman

Amazon Freight is a business unit at Amazon that moves freight for other shippers using Amazon's internal middle mile network. In this talk, we describe the machine learning and price optimization algorithms used to manage Amazon Freight spot market.

For this business, external shippers use a website to request a price quote for a specific origin-destination pair, get prices instantly, and then decide which option to book. Amazon Freight provides prices that varies across options. The execution costs and demand can vary widely across different quotes due to internal cost structure and overall freight market. The goal for Amazon Freight is to set prices for each quote in order to maximize a business objective.

In order to capture the customer's choice of different options with different prices, we employ Multinomial (MNL) choice model which has quote and price dependent features. We use quote features to segment the market into disjoint groups. We employ the Market Segmentation Tree (MST) algorithm, which creates a binary tree based on differences in choice behavior. To optimize prices, we leverage the structure of reference-price-effects MNL choice model to develop a fast two parameter heuristic.

From live A/B experiments, this new framework significantly outperformed the previous system which relied on a mix of machine learning and multi-armed-bandit techniques. The new system increases both profit and revenue and improves prediction accuracy. Moreover, this streamlined approach provides a robust platform to support future services of Amazon Freight.

ME24

Summit - 344

Optimization and Simulation Techniques for Industrial Applications

Contributed Session

Chair: Juhee Kim, Ewha Womans University, Seoul, N/A, Korea, Republic of

1 - Optimal Scheduling Problems with Conflict Constraints: The Wine Fermentation Tanks Case

Alejandro Mac Cawley, Pontificia Universidad Católica de Chile, Santiago, Chile, Susan Cholette, Carlos Monardes, Sergio Maturana, Jorge Vera, Arturo Wenzel

Scheduling and production planning problems are significant for operations across different industries. These problems are challenging, but integer programming (IP) models can solve them. These models support the planner's decision-making to collect and combine the raw material and assign it to some processor during a planning horizon. The time representation into those models can complicate their resolution, especially when the models are time-indexed. This difficulty opens new opportunities to represent the time in the IP models. In this article, we proposed a non-time indexed formulation for a specific type of scheduling problem with the goal of improve the resolution time and handle bigger instances. Our case study is focused on the wine industry. The choices are based on the lot's size, the processors' capacities, the time windows to collect and combine lots, and processing times. We created different instances for this problem to compare the non-time-indexed formulation with another one, which is, in fact, a time-indexed one. We concluded that scheduling problems could be modeled with formulations different from the time-indexed one, which can be more efficient and solve larger scale problems.

2 - Title: Process control optimization with time-series foundation models: advantages, challenges and applications

Markus Ettl, IBM Research, Yorktown Heights, NY, United States, Chandra Reddy, Kanthi Sarpatwar, Wesley Gifford, Vinicius Lima, Dzung Phan, Stuart Siegel, Shivaram Subramanian, Pankaj Dayama, Vijay Ekambaram, Arindam Jati, Binny Samuel, Carlos Hoyos, Gururaj Sagarad, Jayant Kalagnanam

Process-control problems arising in industrial applications have long employed Model Predictive Control (MPC) methods to improve production processes. The estimation methods, for modeling the processes, range from purely physics-based first-principles models to purely data-driven models. With the advances in machine-learning methods for time-series data and the availability of large volumes of sensor data, data-driven methods have gained popularity. Among the data-driven methods, timeseries foundation models (TSFMs), a class of deep-learning methods, especially, have proven to be effective on multi-variate time-series data. However, the deep-learning methods primarily addressed forecasting problems and do not explicitly consider different kinds of variates --- such as control and environment variables. In addition, these models being complex pose problems for control optimization, such as difficulty of obtaining gradients and sensitivities. In this talk, we outline effectiveness of TSFMs for process control, related challenges and solutions. Further, we demonstrate these solutions in the context of control optimization in industries such as cement manufacturing.

3 - Metric Learning-Based Semi-supervised Learning for Machinery Fault Diagnosis

Yunyoung Park, Incheon National University, Incheon, Korea, Republic of, Jaehong Yu

In recent years, deep neural network models have been successfully used for fault diagnosis of the machinery. Although the deep neural network-based models have properly worked under sufficient reliable label information on the operational status, those label information is often considerably insufficient in real situations due to expensive cost of labeling task. To address the label information insufficiency, deep semi-supervised learning methods have been proposed. However, existing deep semi-supervised learning methods might fail to achieve satisfactory performance in the fault diagnosis problem in that these methods suffers from inductive bias issues entailed with pseudo labeling procedure. For more superior fault diagnosis performance, in this study, we newly proposed a metric learning-based semi-supervised fault diagnosis method. The proposed model uses metric learning to extract more appropriate latent feature space where the signals having same label are close together, vice versa for different labels are far apart as possible. By doing so, the pseudo labeling outcomes gets more reliable, and it eventually improve the fault diagnosis performance. We conducted an experimental study using vibration signals obtained from a rolling element bearing simulator, and the results confirmed that the superiority of the proposed framework for detecting rotating machinery faults.

4 - Matrix Representation of the Elementary Operations of Individual Devices and Job Sequence for Control Programming

Hyeon min Yoon, Pohang University of Science and Technology, Pohang, Korea, Republic of, Duck Young Kim

The conventional procedure for manufacturing control programming is to define control logic gates and statements using the pre-specified hardware I/O information for the job sequence corresponding to manufacturing process requirements. Iteratively defining, testing, and adjusting control logic is typically time-consuming, often leading to delays in manufacturing project implementation. Therefore, in this paper, we discuss how to efficiently define, optimize, and verify control logics by automating the control programming process. To achieve this, we first present a mathematical representation in the form of a computer-friendly control matrix of the elementary operations of individual devices, their precedence and logical relationships, and changes in sensor values for each operation. A job sequence can then be represented as a series of matrix operations. We add extra entries to the control matrix as slack variables to calculate the state-transition results following the completion of each individual job. The proposed method is illustrated and validated with control logic programming for a robotized manufacturing workcell.

5 - Learning from Imbalanced Data with SMOTE-like Oversampling for Mixed Type and High Dimensional Data

Juhee Kim, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

The class imbalance problem refers to the number of samples corresponding to each class in the data is not equal, which degrades model performances in machine learning tasks. To overcome this problem, various approaches have been studied, among which data-level resampling methods, including the Synthetic Minority Oversampling Technique (SMOTE) are widely used. The SMOTE and SMOTE-variants are successful oversampling algorithms to obtain balanced data for imbalanced data, but these methods still have difficulty in addressing class imbalance problems from different scenarios, such as the existence of class overlapping regions, high-dimensional data, and data with variables of mixed types. In this paper, we propose a novel oversampling algorithm for mixed type and high-dimensional data which performs re-sampling in a way that appropriately mitigates the problems of imbalanced data. The proposed method consists of four steps: grouping, selection, generation, and validation process. To avoid noisy sample generation in class overlapping region, the proposed method considers not only the information from each of the majority and minority classes, but also the relationship between classes. A new distance metric for both numerical and categorical variables is used throughout the process to allow for sample generation that better reflects the information in both categorical and numerical ones. Additionally, this method includes a verification procedure for the synthesized samples to ensure safe generation.

ME25

Summit - 345

COPT/Responsive.net

Invited Session

Technology Showcase

1 - Unlocking Optimization Potential with COPT

Tiancheng Zhang, Cardinal Operations, Shanghai, China, People's Republic of

Join us for an insightful session on Cardinal Optimizer (COPT), where you'll discover its advanced and newly released capabilities for modeling and solving complex optimization problems. This presentation will introduce the current state of COPT and highlight new features, including a modeling interface for conic problems, support for the exponential cone, and first-order methods on GPU. Initial results on solving large-scale LPs on GPU show exciting possibilities to tackle previously unsolvable problems.

As the R&D team continues to achieve technical breakthroughs, our operations and support team are dedicated in engaging and enriching the COPT user community. With its user-friendly interfaces and comprehensive support, COPT has been seamlessly integrated into various packages and software for both research projects and industrial applications.

2 - Littlefield 2.0: A New Version of the Online Game for Operations Management Courses

Samuel Wood, Responsive Learning Techologies, Los Altos, CA, United States

After 25 years there is a new version of Littlefield! Littlefield is a competitive online simulation of either a factory or a medical laboratory that has been by more than half a million students in 500+ universities in 60+ countries to excite and engage students in operations management topics like process analysis and inventory control. This presentation will introduce a newly updated version 2 of the game that was released this past summer.

ME26

Summit - 346

Advanced Bandit Algorithms and Optimization Techniques

Contributed Session

Chair: Kemal Gursoy, Rutgers University, 100 Rockafeller Road, Dept of MSIS, Room 5146, Piscataway, NJ, 08854, United States

1 - Multi-Armed Bandits and Games

Kemal Gursoy, Rutgers University, Piscataway, NJ, United States

A sequential optimization model, the multi-armed bandit problem, is concerned with optimal allocation of resources between competing activities, to generate the most likely benefits. In this work, following the objective of a multi-armed bandit problem, we consider a mean-field game theoretic model to approach to a large number of multi-armed bandits, and propose some connections between dynamic games and sequential optimization problems.

2 - A Multi-Task Deep Learning Approach to Large-Scale Lifetime Value Evaluation in Mobile Gaming Market

Qingpeng Zhang, The University of Hong Kong, Kowloon, Hong Kong, Mengzhuo Guo, Yijun Li, Binfeng Jia, Yang Liu

In the growing mobile gaming market, monetization tactics that encourage high-value transactions depend on the accurate and granular evaluation of user lifetime value (LTV). Existing probabilistic methods cannot handle multiple types of high-dimensional data with heavy-tailed distributions of in-game purchase amounts. Existing deep learning methods, on the other hand, do not account for the user heterogeneity across different mobile games. This study proposes an end-to-end deep multi-task learning approach for LTV prediction based on the Fogg Behavior model (FBM). It models the user's motivation and ability to pay as the probability and expectation of payment amount, respectively, and considers the co-effect of multiple mobile games. Besides achieving state-of-the-art performance in offline experiments, it also provides explainable results for analyzing user heterogeneity by examining their motivation and ability. We conducted large-scale field experiments on a real online advertising platform, showing that the proposed approach increased advertisers' Return on Investment by an average of 0.469\% for every 1\% improvement in the LTV prediction performance. Our approach is now deployed online, impacting billions of traffic daily.

3 - DISTRIBUTIONALLY ROBUST TWIN SUPPORT VECTOR MACHINE

Lesheng Wang, North Carolina State University, Raleigh, NC, United States, Fengming Lin, Shu-Cherng Fang, Xinyu Xu

In this presentation, we propose a distributionally robust twin support vector machine model to address the classification problems with uncertainty using the first and second-order information. The proposed model finds two nonparallel linear separation hyperplanes by solving two interconnected chance-constrained SVM models. Tractable SDP and SOCP reformulations are derived for efficient computations. Computational results on synthetic data and real-world benchmarks showcase superior performance of the proposed model compared with other established classification models, particularly for imbalanced datasets.

4 - Critical Analysis on Bio Inspired Optimization Methods and its Applications

Amit Kumar Saha, University of Arkansas, Fayetteville, AR, United States, Haitao Liao

In contemporary times, optimization techniques have evolved significantly in computing and complex problem-solving. Real-world problem formulation requires tuning various system parameters to achieve the best results within limited resources using diverse optimization methods. Traditional optimization techniques often rely on specific mathematical approaches tailored to applications. Recently, bio-inspired optimization has gained popularity due to its foundation in the practical behaviors of biological systems, which inherently seek optimal solutions to maintain systemic balance. When the design surface is non-linear, these optimization methods are very effective to finding the best solution. Numerous studies over the past few years have explored techniques that emulate these biological behaviors. This paper reviews the development, classification, performance and functioning of several prominent bio-inspired methods, highlighting their applications to real-world problems. It also identifies the gaps between theoretical foundations and practical applications of these techniques, suggesting directions for future research.

ME27

Summit - 347

TIMES Distinguished Speaker Award Session

Award Session

Technology, Innovation Management and Entrepreneurship

Chair: Gizem Korpeoglu, TU/e, Eindhoven, 5632TE, Netherlands

1 - Jurgen's Talk

2 - Discussion of Key Topics in Innovation

Stylianos Kavadias, University of Cambridge, Cambrige, United Kingdom

Invited discussion of my perspective of the Innovation works that I have contributed to over the past 25 years.

ME28

Summit - 348

Advances in Causal Inference and Predictive Analytics

Contributed Session

Chair: Hossein Tohidi, SAS, 4131 Briargrove Circle, Apt 403, Raleigh, NC, 27607, United States

1 - Conformal Prediction for Interval Outcomes

Weiguang Liu, UCL, London, United Kingdom, Aureo de Paula, Elie Tamer

Interval censoring often occurs in economic data, leading to partial identification. Predicting outcomes that are interval censored is important in many practical scenario. We propose a new method to construct the prediction intervals for interval censored outcomes, by leveraging a natural characterisation of the sharp identified region of conditional cumulative distribution functions (CDF) rather than the conditional mean and the method of conformal inference. The proposed method is shown to both have finite-sample coverage guarantees and achieves asymptotic efficiency in terms of the volume of the prediction set.

2 - Confusion with the Confusion Matrix: a Meta-Analysis on Threshold Setting in Binary Classification

Patrick Moder, Kühne Logistics University, Hamburg, Germany, Kai Hoberg

Studies on binary classification usually report the balance of expected positive and negative outcomes (both true and false) to display classification model performance. Varying the decision threshold, which determines the cutoff between positive and negative classifications, calibrates that balance to maximize performance or minimize costs for expected misclassifications. In operations management, an optimal balance depends on study context and misclassification costs. The consequences of false alarms (false positive) can be less severe (costly) than missing a target (false negative), thus a certain balance of classifications is optimal. Yet, several studies do not reveal the rationale for calibrating classification outcomes. Hence, it is difficult to understand whether the results show a cost- or performance-optimizing solution, or an arbitrary (default) threshold was selected. When studies present their classification results in a confusion matrix or with relative metrics (precision and recall, e.g.), only outcomes for one particular decision threshold are shown. It is thus unclear how the performance varies at different thresholds and models can be operationalized. We conduct a meta-analysis on a large sample of research articles in operations management that present results of binary classifications and unveil how classifications are balanced. We describe used strategies to calibrate classifications, provide explanations for biased and boundedly rational calibration (anticipated regret or cry-wolf effect, e.g.), and provide recommendations for a rational and transparent reporting of classification performance. Our results have implications for researchers and practitioners who intend to rigorously report (with threshold-independent performance metrics, e.g.) or consciously interpret binary classification results.

3 - A Mixed Integer Programming Model for Designing Causal Inference Networks

Ali Ilhan Haliloglu, Bilkent University, Ankara, Turkey, Oya Ekin Karasan, Ozlem Karsu

In the realm of causal inference and probabilistic reasoning, understanding the relationships between variables is paramount. D-Separation and Connection, two fundamental concepts in the field of graphical models, provide a structured framework to determine whether variables are independent given a set of observed variables. D-Separation helps identify when variables are independent given certain observed variables, aiding in simplifying complex models. On the other hand, Connection highlights scenarios where variables are dependent, even when there is no direct link between them. Together, D-Separation and Connection offer invaluable insights into untangling the intricate web of causal relationships, enabling better decision-making and inference in diverse disciplines. Our research aims to write a mathematical model that will yield a graphical network with minimum weighted violations between dependent and independent variables with given collider sets. The weights of the relationships are determined with a known BIC test from the literature. The model uses a mixture of labeling and flow based modeling approach where variables are represented as nodes and edges represents the relationships between these variables. The model takes the dependence and independence relationships as parameters and the type of edges as decision variables and yields a network with minimum amount of weighted violation between variables.

4 - Graph-Based LogP Prediction: Leveraging Molecular Structure with Graph Neural Networks

Hossein Tohidi, SAS, Cary, NC, United States, Brandon Reese

Abstract:

Molecular structures can be effectively represented as graphs, with atoms as nodes and bonds as edges, offering rich connectivity patterns crucial for property prediction tasks. LogP, the logarithm of the partition coefficient, serves as a pivotal descriptor reflecting a molecule's hydrophobicity, influencing its behavior in biological processes and chemical environments. In this study, we propose a novel supervised approach that employs feature engineering and Graph Neural Networks (GNNs) to accurately predict LogP values from molecular graphs.

Our method exploits the inherent graph representation of molecules to systematically extract informative node- and graph-level features crucial for LogP prediction. Leveraging SAS Viya capabilities as well as the PyTorch Geometric Python package, we train our model on diverse datasets, ensuring robustness and generalization across various molecular structures. We conduct extensive experiments, benchmarking our approach against state-of-the-art methods in LogP prediction. Our results underscore the efficacy of our graph-based model in accurately predicting LogP values.

In summary, our work showcases the potential of Graph Neural Networks in elucidating graph properties for LogP prediction. By integrating graph representation learning with molecular property prediction, we offer a promising framework to enhance drug discovery and molecular design processes.

5 - Robust Bayesian Inference of Causal Effects via Randomization Distributions

Easton Huch, University of Michigan, Department of Statistics, Ann Arbor, MI, United States, Walter Dempsey, Fred Feinberg

We present a general framework for Bayesian inference of causal effects that delivers provably robust inferences based principally on the physical act of randomization. The framework involves fixing the observed potential outcomes and forming a simulation-based likelihood via the randomization distribution of a model-based discrepancy variable, a summary of the (imputed) complete data. We show posterior consistency of the method and derive theoretical connections to common estimators in causal inference. We illustrate the method and theoretical results in simulation and a case study.

ME29

Summit - 420

Intersection of Machine Learning and Integer Programming

Invited Session

OPT: Integer and Discrete Optimization

Chair: Tengmu Hu, The Ohio State University, Columbus, OH, 43202, United States

Co-Chair: Tengmu Hu, The Ohio State University, Columbus, OH, 43202, United States

1 - An Integer Column Generation Algorithm for the Large Scale Optimal Classification Tree Problem

Tengmu Hu, The Ohio State University, Columbus, OH, United States

This paper explores the application of MIP techniques to enhance the efficiency of solving optimal classification tree problems. We introduce an exact integer column generation (ICG) algorithm and reformulate the problem as a set packing problem, leveraging its structural properties. Additionally, we investigate the convex hull of optimal classification tree formulations under tree depth equals to one. Both proposed pricing subproblems can be efficiently solved in polynomial time. Furthermore, we propose strategies for column addition and the generation of the next integer solution that yield maximal improvement.

2 - Learning to Disable Globally-Valid Cuts

Zixuan Feng, University of Florida, Gainesville, FL, United States

Cutting planes, or cuts, are a relaxation tightening technique employed by Mixed-Integer Programming (MIP) solvers, due to the fact it improves the average solving time. However, this is not always the case. For some instances, adding cuts can actually slow down the solving process for reasons including but not limited to the inequalities created are too dense, or too many cuts are added. Our research focuses on the Branch-and-Cut (B&C) method and aims to predict whether a MIP problem will benefit from disabling globally-valid cuts with the help of Machine Learning (ML) methods. Additionally, we tested different timings for making such a decision. Our study is tested with the MIPLIB 2017 benchmark dataset and showed promising results on improving the solving time by making adaptive choices.

3 - Learning to Branch in AC-Network Constrained Unit Commitment Problems

Selin Bayramoglu, Georgia Institute of Technology, Atlanta, GA, United States, George Nemhauser, Nikolaos Sahinidis

Unit commitment (UC) problems are challenging problems that power system operators solve multiple times a day. These problems are similar, differing only in some of the problem data involved, such as demands (loads). This work develops a machine-learning methodology to build a model that predicts strong branching scores—a costly expert branching rule—to speed up the solution of UC problems. The UC formulation in this study contains nonconvex AC power flow constraints and is thus a nonconvex mixed-integer nonlinear programming problem. This work shows that the SCIP solver equipped with the learned branching rule is faster than its default.

4 - Sample Complexity of Learning Cut Generating Functions

Hongyu Cheng, Johns Hopkins University, Baltimore, MD, United States, Amitabh Basu

Selecting an effective cutting plane to minimize the branch-and-cut tree size is a critical challenge in the branch-and-cut algorithm. Recent advancements have employed a data-driven approach to select optimal cutting planes from a parameterized family, aimed at reducing the branch-and-bound tree size for a given distribution of integer programming instances. We extend this idea to the selection of the best cut generating function (CGF) from a parameterized family of facet-defining valid functions for the infinite group relaxation problem, with Gomory's mixed-integer (GMI) cuts as a particular instance. We provide rigorous sample complexity bounds for the selection of an effective CGF from these families for any specified distribution under some mild assumptions. Our empirical results show that the selected CGF can outperform the GMI cuts for certain synthetic distributions. Additionally, we explore the sample complexity of using neural networks for instance-dependent CGF selection.

ME30

Summit - 421

ONR Optimization Program Review: Applied Models and Algorithms

Invited Session

OPT: Computational Optimization and Software

Chair: David Phillips, Office of Naval Research, University of New Mexico, Catonsville, MD, United States

1 - New Perspectives on Critical Elements Detection in Complex Networks

Jourdain Lamperski, University of Pittsburgh, Pittsburgh, PA, United States, Oleg Prokopyev

Assessing a complex network’s vulnerability and quantifying its robustness properties with respect to adversarial attacks, natural disasters, and random failures is an important research question that arises in a variety of applications in telecommunication, transportation, social science, energy, and homeland security. This research challenge is often addressed using the graph-theoretic notion of critical elements, namely a subset of elements in a graph that satisfy some budget and/or structural constraints and whose removal (by some decision-maker) maximally degrades some desirable properties of the graph (e.g., its connectivity). Existing models typically use overly simplistic graph-theoretic notions as objective functions (e.g., the number of connected node pairs). The goal of part of our project is to consider more complex objective functions that build upon a rich body of related work in the network analysis literature. In this talk we will summarize results in this direction, specifically related to algebraic connectivity, a spectral graph measure that is often used to study robustness and synchronization properties of networks. The primary focus will be on approximation algorithms, but we will also discuss hardness results as well as exact algorithms based on combinatorial optimization formulations.

2 - Stochastic Programming Under Endogenous Uncertainty: Electric Vehicle Fleet for Energy Network Resilience

Miguel Lejeune, George Washington University, Washington, DC, United States, Payman Dehghanian

We propose new chance-constrained optimization models with decision-dependent uncertainty for the routing and scheduling of a fleet of electric vehicles used for enhancing energy network resilience and restoring service. Decisions pertaining to the assignment, dispatch, and coordination of the electric vehicle fleet and repair crews must be taken. The proposed models account for the coupling and interdependence between energy (power transmission and distribution) and transportation (ground and railway) networks under exogenous and endogenous uncertainties. We develop a solution approach that interlaces several reformulation and algorithmic methods. In general, the reformulated problems take the form of non-convex mixed-integer nonlinear problems. The algorithmic framework features optimality-based bound tightening methods and a specialized nonlinear branch-and-bound.

3 - An Overview of Optimizing Network Navigation in Stochastic Obstacle Scene Problem

Elvan Ceyhan, Auburn University, Auburn, AL, United States

This research focuses on optimizing navigation in networks with obstacles to address the stochastic obstacle scene (SOS) problem. This study examines two variants: the Optimal Traversal Path (OTP) problem and the Optimal Obstacle Placement (OOP) problem. The OTP problem involves a single navigating agent (NAVA) selecting a cost-minimal path through a space containing ‘forbidden region’. The OOP problem, recently introduced by the investigator, involves an obstacle-placing agent (OPA) strategically inserting obstacles to maximize the traversal length of the NAVA. The main goals are to improve heuristic algorithms for both SOS variants, introduce new strategies, extend the SOS problem to higher dimensions, develop weight-constrained versions of both variants, and analyze the theoretical properties and complexities of the algorithms. The application areas include defense logistics, autonomous path planning, target tracking, and supply chain management.

Keywords: Canadian traveler’s problem, spatial randomness, spatial pattern, network traversal and blocking, heuristic algorithm; stochastic ordering, adversarial agents, algorithm complexity

Joint work with Polat Charyyev, MAP Akademi, Istanbul, Turkey and Li Zhou, Auburn University, AL, USA

4 - Navy Vehicle Scheduling: Problem Formulation and Solutions

Sungjin Im, University of California, Merced, Merced, CA, United States, Mugen Blue, Benjamin Moseley, Rudy Zhou

The Navy operates a large number of vehicles. Scheduling these deployments presents a complex problem due to the critical nature of both the valuable hardware and the crew. These challenges are further exacerbated by various non-standard and non-trivial constraints. For example, vehicle operations must be accompanied by sufficient homeport training to ensure both deployment readiness and the long-term well-being of the crew.

In this talk, we will present the scheduling problems and challenges faced by the Navy, discuss how we model them, and explore how our model addresses these challenges.

ME31

Summit - 422

Large Language Models and Optimization

Invited Session

OPT: Machine Learning

Chair: Can Li, Purdue University, West Lafayette, IN, United States

1 - Ai and the Future of Optimization Modeling

Madeleine Udell, Stanford University, Stanford, CA, United States, Ali AhmadiTeshnizi, Wenzhi Gao

Optimization problems are pervasive in sectors from manufacturing and distribution to healthcare. However, most such problems are still solved heuristically by hand rather than optimally by state-of-the-art solvers, as the expertise required to formulate and solve these problems limits the widespread adoption of optimization tools and techniques.

As a glimpse of the future, this talk will introduce OptiMUS, a Large Language Model (LLM)-based agent designed to formulate and solve MILP problems from natural language descriptions. OptiMUS can develop mathematical models, write and debug solver code, develop tests, and check the validity of generated solutions. Experimentally, OptiMUS correctly solves more than 80% of benchmark problems, more than twice as many as a basic LLM prompting strategy. More broadly, we discuss the potential for LLMs in domains where accuracy and fidelity to real-world data is critical and strategies to augment and safeguard their performance.

2 - Synthesizing Mixed-Integer Linear Programming Models from Natural Language Descriptions

Qingyang Li, The University of Melbourne, Melbourne, Australia, Lele Zhang, Vicky Mak-Hau

Numerous real-life problems arising in various domains can be formulated and solved using Mixed-Integer Linear Programming (MILP) models. However, the transformation of real-world decision-making problems into MILP models requires expertise in Operations Research and mathematical optimization, which restricts non-experts' accessibility to MILP. To address this challenge, we propose a framework for automatically formulating MILP models from unstructured natural language descriptions of decision problems, which integrates Large Language Models (LLMs) and mathematical modeling techniques. This framework consists of three phases: i) identification of decision variables, ii) classification of objective and constraints, and iii) generation of MILP models.

In this study, we propose a knowledge representation structure that includes relational knowledge and inheritable knowledge. Specifically, we present a constraint classification scheme and a set of constraint templates that can guide the LLMs in synthesizing a complete MILP model from natural language. Our approach can synthesize logic constraints and classic demand and resource constraints.

To assess our framework, we extend the NL4Opt dataset with more constraint types, and create a new test dataset. With respect to the accuracies of generating the correct model, objective, and constraints, our method which integrates knowledge representation with LLMs significantly outperforms one-step model generation methods offered by ChatGPT and Google Bard. Our three-phase prototype system for automatically formulating MILP models has a great potential to capture more types of constraints for more complex problems. It opens up opportunities for developing training tools for operations research practitioners and powerful tools for automatic decision problem modeling and solving in practice.

3 - "I Want It that Way": Enabling Interactive Decision Support via Large Language Models and Constraint Programming

Connor Lawless, Cornell University, Ithaca, NY, United States

Mathematical Programming has been a transformative technology for improving decision making in a number of industrial applications but its impact has been largely realized by large institutions that can afford the optimization expertise. We present a novel framework for interactive decision support for non-expert users that leverages large language models (LLM) to translate natural language requests into operations on an underlying constraint programming model. We study this hybrid framework through the lens of meeting scheduling, a time-consuming daily activity faced by a multitude of information workers. We conduct three studies to evaluate the novel framework, including a diary study (n=64) to characterize contextual scheduling preferences, a quantitative evaluation of the system's performance, and a user study (n=10) with a prototype system. Our work highlights the potential for a hybrid LLM and optimization approach for iterative preference elicitation and design considerations for building systems that support human-system collaborative decision-making processes.

4 - Explaining Optimization Models Using LLMs

Can Li, Purdue University, West Lafayette, IN, United States, Hao Chen, Gonzalo Constante Flores

We will present OptiChat, a chatbot for explaining optimization models through sensitivity analysis, infeasible diagnosis, and counterfactual explanation that is powerd by large language models.

ME32

Summit - 423

Robust Optimization for Social Good

Invited Session

OPT: Optimization Under Uncertainty

Chair: Qing Jin, USC University of Southern California, LA, CA, United States

Co-Chair: Phebe Vayanos, University of Southern California, Los Angeles, CA, United States

1 - Robust Offline Policy Learning Under Covariate Shift

Bill Tang, University of Southern California, Los Angeles, CA, United States, Cagil Kocyigit, Phebe Vayanos

We study the problem of distribution shifts in offline policy learning, where
the policy training distribution is different from the deployment distribution and may lead to harmful or sub-optimal policy actions at deployment. In real world applications, changes to an allocation system can cause shifts in measured covariates, such as wording changes in survey questions that elicit different responses from individuals experiencing homelessness. As a result, a non-robust allocation policy may incorrectly over or under allocate resources based on the original offline data distribution. Motivated by this, we seek to learn an allocation policy that is not restricted to any functional form and robust to potential covariate shifts in the population of allocatees.

2 - Wasserstein Logistic Regression Robust to Distribution Shifts

Qingshi Sun, University of Southern California, Los Angeles, CA, United States, Nathan Justin, Andres Gomez, Phebe Vayanos

Logistic regression is one of the most popular classification models in high-stakes domains due to its great interpretability. Also, in high-stakes domains, data tend to be collected by surveys that are sensitive to many factors, such as the level of comfort respondents have in providing information, when and where the survey is conducted. These factors may affect the data collection and lead to shifts in data distribution. Therefore, we propose a distributionally robust logistic regression model robust to distribution shifts. Compared to recent work, our model includes all kinds of features (i.e., numerical, integer, and categorical) and thus significantly extends the fields of application. Also, the model includes the likelihood of the perturbation of each feature. We evaluate the performance of our model on numerous publicly available datasets, and compare the performance to a regularized logistic regression. Our proposed model achieves improvement in calibration error and the area under the ROC curve (AUC). Despite the distributionally robust model amounting to an exponential-size optimization problem, we reformulate the model equivalently by combining the shortest path problem based on dynamic programming digraphs. This allows the reformulated model to be solved by an off-the-shelf solver in polynomial time.

3 - Learning Optimal Classification Trees Robust to Distribution Shifts

Nathan Justin, University of Southern California, Los Angeles, CA, United States, sina Aghaei, Andres Gomez, Phebe Vayanos

We consider the problem of learning classification trees that are robust to distribution shifts between training and testing/deployment data. This problem arises frequently in high stakes settings such as public health and social work where data is often collected using self-reported surveys which are highly sensitive to e.g., the framing of the questions, the time when and place where the survey is conducted, and the level of comfort the interviewee has in sharing information with the interviewer. We propose a method for learning optimal robust classification trees based on mixed-integer robust optimization technology. In particular, we demonstrate that the problem of learning an optimal robust tree can be cast as a single-stage mixed-integer robust optimization problem with a highly nonlinear and discontinuous objective. We reformulate this problem equivalently as a two-stage linear robust optimization problem for which we devise a tailored solution procedure based on constraint generation. We evaluate the performance of our approach on numerous publicly available datasets, and compare the performance to a regularized, non-robust optimal tree. We show an increase of up to 12.48% in worst-case accuracy and of up to 4.85% in average-case accuracy across several datasets and distribution shifts from using our robust solution in comparison to the non-robust one.

4 - Biodiversity Conservation via Adjustable Robust Optimization

Qing Jin, University of Southern California, LA, CA, United States, Yingxiao Ye, Christopher Doehring, Angelos Georghiou, Hugh Robinson, Phebe Vayanos

Human development is a threat to biodiversity and conservation organizations (COs) are purchasing land to protect areas for biodiversity preservation. COs have limited budgets and cannot purchase all the land necessary to perfectly preserve biodiversity, and human activities are uncertain, so exact developments are unpredictable. We propose a multistage, robust optimization problem with a data-driven hierarchical-structured uncertainty set which captures the endogenous nature of the binary (0-1) human land use uncertain parameters to help COs choose land parcels to purchase to minimize human impact on biodiversity. We simulate the development based on climate data, land characteristics, and human land use data. We use the simulation to build data-driven uncertainty sets. We demonstrate that an equivalent formulation of the problem can be obtained that presents exogenous uncertainty only and where uncertain parameters only appear in the objective. We leverage this reformulation to propose a conservative approximation and showcase the proposed approach outperform the existing methods widely used in the biodiversity conservation community.

5 - Mistake, Manipulation and Margin Guarantees in Online Strategic Classification

Lingqing Shen, Carnegie Mellon University, Pittsburgh, PA, United States, Nam Ho-Nguyen, Khanh-Hung Giang-Tran, Fatma Kilinc-Karzan

We consider an online strategic classification problem where each arriving agent can manipulate their true feature vector to obtain a positive predicted label, while incurring a cost that depends on the amount of manipulation. The learner seeks to predict the agent's true label given access to only the manipulated features. After the learner releases their prediction, the agent's true label is revealed. Previous algorithms such as the strategic perceptron guarantee finitely many mistakes under a margin assumption on agents' true feature vectors. However, these are not guaranteed to encourage agents to be truthful. Promoting truthfulness is intimately linked to obtaining adequate margin on the predictions, thus we provide two new algorithms aimed at recovering the maximum margin classifier in the presence of strategic agent behavior. We prove convergence, finite mistake and finite manipulation guarantees for a variety of agent cost structures. We also provide generalized versions of the strategic perceptron with mistake guarantees for different costs. Our numerical study on real and synthetic data demonstrates that the new algorithms outperform previous ones in terms of margin, number of manipulation and number of mistakes.

ME33

Summit - 424

Algorithms for Robust Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Kartikey Sharma, Zuse Institute Berlin, Berlin, Germany

1 - Tractable Optimization Under Multilinear Uncertainty

Kartikey Sharma, Zuse Institute Berlin, Berlin, Germany, Eojin Han, Omid Nohadani

Capturing the endogeneity of uncertainity can lead to non-linear models. However, most research on optimization under uncertainty has focused on linear settings. We develop reformulation techniques for Robust and Distributionally Robust Optimization to tackle uncertainties which affect the optimization problem through a multilinear function. For RO, we provide a general conservative reformulation and provide situations in which the reformulation is exact whereas for DRO we show that for particular ambiguity sets the problem can be reformulated as a stochastic program. These techniques lead to efficient algorithms to solve problems with recursive uncertainties such as balking.

2 - A Nonparametric Approach with Marginals for Modeling Consumer Choice

Yanqiu Ruan, Singapore University of Technology and Design, Singapore, Singapore, Xiaobo Li, Karthyek Murthy, Karthik Natarajan

Given data on the choices made by consumers for different offer sets, a key challenge is to develop parsimonious models that describe and predict consumer choice behavior while being amenable to prescriptive tasks such as pricing and assortment optimization. The marginal distribution model (MDM) is one such model, which requires only the specification of marginal distributions of the random utilities. This paper aims to establish necessary and sufficient conditions for given choice data to be consistent with the MDM hypothesis, inspired by the utility of similar characterizations for the random utility model (RUM). This endeavor leads to an exact characterization of the set of choice probabilities that the MDM can represent. Verifying the consistency of choice data with this characterization is equivalent to solving a polynomial-sized linear program. Since the analogous verification task for RUM is computationally intractable and neither of these models subsumes the other, MDM is helpful in striking a balance between tractability and representational power. The characterization is convenient to be used with robust optimization for making data-driven sales and revenue predictions for new unseen assortments. When the choice data lacks consistency with the MDM hypothesis, finding the best-fitting MDM choice probabilities reduces to solving a mixed integer convex program. Numerical experiments show that MDM provides better representational power and prediction accuracy than multinominal logit and significantly better computational performance than RUM.

3 - Primal-Dual Algorithms for Stochastic Distributionally Robust Optimization

Zaid Harchaoui, University of Washington, Seattle, WA, United States

We consider the penalized distributionally robust optimization (DRO) problem with a closed, convex uncertainty set, a setting that encompasses the f-DRO, Wasserstein-DRO, and spectral/L-risk formulations used in practice. We present Drago, a stochastic primal-dual algorithm that achieves a state-of-the-art linear convergence rate on strongly convex-strongly concave DRO problems. The method combines both randomized and cyclic components with mini-batching, which effectively handles the unique asymmetric nature of the primal and dual problems in DRO. We support our theoretical results with numerical benchmarks in classification and regression. This is joint work with Ronak Mehta and Jelena Diakonikolas.

4 - Distributed Robust Optimization for Multi-Agent Robotic Systems

Evangelos Theodorou, Georgia Institute of Technology, Atlanta, GA, United States

For multi-agent autonomous robotic systems to successfully perform challenging tasks, they must robustly operate in dynamic and uncertain environments. While there has been a plethora of prior work on distributed decision-making, there are no scalable algorithms that can handle different types of uncertainty. Decision-making algorithms based on robust optimization have the capability to handle deterministic and stochastic uncertainties. However, this capability comes at the price of high computational cost which prohibits the applicability of robust optimization to large-scale multi-agent robotics systems. In this talk we present a scalable distributed robust optimization framework that addresses these issues. The proposed approach can handle different types of uncertainties while retains computational efficiency. The presentation includes tractable approximations of robust constraints that are relevant to multi-robot settings as well as a complexity analysis. Several examples and use cases on multi-agent robotic navigation tasks will be presented to demonstrate the scalability and efficacy of the proposed distributed robust trajectory optimization framework. Finally, future directions on the theory and applications will be discussed.

ME34

Summit - 425

First-order Methods in Continuous and Stochastic Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Jiaming Liang, University of Rochester, Rochester, NY, United States

1 - a Simple Uniformly Optimal Method Without Line Search for Convex Optimization

Tianjiao Li, Georgia Institute of Technology, Atlanta, GA, United States, Guanghui Lan

Line search (or backtracking) procedures have been widely employed into first-order methods for solving convex optimization problems, especially those with unknown problem parameters (e.g., Lipschitz constant). In this paper, we show that line search is superfluous in attaining the optimal rate of convergence for solving a convex optimization problem whose parameters are not given a priori. In particular, we present a novel accelerated gradient descent type algorithm called auto-conditioned fast gradient method (AC-FGM) that can achieve an optimal O(1/k^2) rate of convergence for smooth convex optimization without requiring the estimate of a global Lipschitz constant or the employment of line search procedures. We then extend AC-FGM to solve convex optimization problems with Holder continuous gradients and show that it automatically achieves the optimal rates of convergence uniformly for all problem classes with the desired accuracy of the solution as the only input. Finally, we report some encouraging numerical results that demonstrate the advantages of AC-FGM over the previously developed parameter-free methods for convex optimization.

2 - Tangent Subspace Descent via Discontinuous Subspace Selections on Fixed-Rank Manifolds

David Gutman, Texas A&M University, College Station, TX, United States, Nam Ho-Nguyen, Daniel Grady

The tangent subspace descent method (TSD) extends the coordinate descent algorithm to manifold domains. The key insight underlying TSD is to draw an analogy between coordinate blocks in Euclidean space and tangent subspaces of a manifold. The core principle behind ensuring convergence of TSD for smooth functions is the appropriate choice of subspace at each iteration. Previously, it was shown that it is always possible to appropriately pick such subspaces on the broad class of manifolds known as naturally reductive homogeneous spaces. In this talk, we provide the first instances of TSD for manifolds outside of this class. The main idea underlying these new instances is the use of discontinuous subspace selections. As a result of our developments we derive new and efficient methods for large-scale optimization on the fixed-rank and fixed-rank positive semidefinite matrix manifolds.

3 - A Nearly Linearly Convergent First-Order Method for "Typical" Nonsmooth Functions

Liwei Jiang, Georgia Institute of Technology, Atlanta, GA, United States

Nonsmooth optimization problems appear throughout machine learning and signal processing. Standard gradient methods in nonsmooth optimization are often described as “slow” since the well-known “lower-complexity bounds” suggest they converge at best sublinearly. In this talk, I will introduce a first-order method that (locally) exponentially improves upon such lower bounds. The method is “parameter-free” and converges nearly linearly on “typical” optimization problems. The key insight is that “typical” nonsmooth functions are not pathological but are “partially smooth” in algorithmically useful ways.

4 - Accelerated Optimization over Smooth and/OR Strongly Convex Sets

Benjamin Grimmer, Johns Hopkins University, Baltimore, MD, United States

We consider feasibility and constrained optimization problems defined over smooth and/or strongly convex sets. These notions mirror their popular function counterparts but are much less explored in the first-order optimization literature. In this talk, we will aim to lessen this gap.

We first extend the computer-assisted performance estimation problem machinery to such problems. From this, we derive new and improved theory for classic methods like Frank-Wolfe and alternating projections directly benefiting from smoothness and strong convexity of constraint sets. Considering new gauge-based methods, we derive optimal accelerated convergence guarantees of O(1/T) for problems with strongly convex constraint sets and O(1/T^2) for smooth constraint sets. This second analysis are based on novel characterizations of the Minkowski gauge of smooth and/or strongly convex sets, which may be of independent interest: although the gauge is neither smooth nor strongly convex, we show the gauge squared inherits any structure present in the set.

ME35

Summit - 427

TSL Best Student Paper Award

Award Session

Transportation Science and Logistics (TSL)

Chair: Mathias Klapp, Pontificia Universidad Católica de Chile, Santiago, Chile

1 - Machine Learning-Augmented Optimization of Large Bilevel and Two-Stage Stochastic Programs: Application to Cycling Network Design

Bo Lin, University of Toronto, North York, ON, Canada

Motivated by a cycling infrastructure planning application, we present a machine learning approach to solving bilevel programs with a large number of independent followers, which as a special case includes two-stage stochastic programming. We propose an optimization model that explicitly considers a sampled subset of followers and exploits a machine learning model to estimate the objective values of unsampled followers. Unlike existing approaches, we embed machine learning model training into the optimization problem, which allows us to employ follower features that cannot be represented using leader decisions. We prove bounds on the optimality gap of the generated leader decision as measured by the original objective that considers the full follower set. We develop follower sampling algorithms to tighten the bounds and a representation learning approach to learn follower features, which are used as inputs to our machine learning model. Through numerical studies, we show that our approach generates leader decisions of higher quality compared to baselines. Finally, we perform a real-world case study in Toronto, Canada, where we solve a cycling network design problem with over one million followers. Compared to the current practice, our approach improves a transportation metric by 19.2% and can lead to a potential cost saving of $18M.

2 - Balancing Fairness and Efficiency in Traffic Routing via Interpolated Traffic Assignment

Devansh Jalota, Stanford University, Stanford, CA, United States

System optimum (SO) routing, wherein the total travel time of all users is minimized, is a holy grail for transportation authorities. However, SO routing may discriminate against users who incur much larger travel times than others to achieve high system efficiency, i.e., low total travel times. To address the inherent unfairness of SO routing, we study a fairness constrained traffic assignment problem which minimizes the total travel time while achieving a desired fairness level specified by the maximum ratio between the travel times of users with shared origins and destinations. To obtain feasible solutions to this problem while achieving high system efficiency, we develop a new convex program, the interpolated traffic assignment problem (I-TAP), which interpolates between a fairness-promoting and an efficiency-promoting traffic-assignment objective. We evaluate the efficacy of I-TAP through theoretical bounds on the total system travel time and level of fairness in terms of its interpolation parameter and present a numerical comparison between I-TAP and a state-of-the-art algorithm. The numerical results indicate that our approach is faster by several orders of magnitude compared to the benchmark, while achieving higher system efficiency for all desirable fairness levels. We further leverage I-TAP's structure to develop pricing mechanisms to collectively enforce the I-TAP solution in the presence of selfish users that independently choose routes to minimize their travel costs. We mention that this is the first study of pricing in the context of fair routing for general road networks (as opposed to, e.g., parallel networks).

3 - Fair Fares for Vehicle Sharing Systems

Hyemi Kim, Columbia University, New York, NY, United States

Vehicle sharing systems, like those for bicycles, scooters, and cars, have become essential transportation options, with fares often set by companies through mobile apps. This can result in users in different locations paying varying prices and experiencing unequal access. Platforms and regulatory bodies should consider how to set prices fairly to minimize the inequalities experienced by users across locations, for which we provide a framework and insights in this work. We consider two notions of fairness corresponding to price and access. Price fairness is a measure of how close the prices at two locations are to one another. Access fairness compares the fraction of demand at each location that has access to the system, where access is a product of affordability and vehicle availability. We analyze the impact of imposing these fairness measures on the revenue, consumer surplus, and social welfare using a stylized model with two locations. Under price fairness, we analytically identify regimes where consumer surplus at both locations can increase or can both decrease. We show that access fairness decreases consumer surplus at both locations, making all parties worse off. To address the non-convexity in generalized networks with multiple locations and vehicles, we design a sliding window method for price fairness and employ a convex relaxation technique achieving asymptotic optimality for access fairness. Lastly, a case study using real-word data of a vehicle sharing system operating in New York City shows that a significant gain in fairness can be achieved with a relatively small loss in revenue

4 - Subpath-Based Column Generation for the Electric Routing-Scheduling Problem

Sean Lo, Massachusetts Institute of Technology, Cambridge, MA, United States

Motivated by widespread electrification targets, this paper studies an electric routing-scheduling problem (ERSP) that jointly optimizes routing-scheduling and charging decisions. The ERSP is formulated as a semi-infinite set-partitioning model, where continuous charging decisions result in infinitely-many path-based variables. To solve it, we develop a column generation algorithm with a bi-level label-setting algorithm to decompose the pricing problem into (i) a first-level procedure to generate subpaths between charging stations, and (ii) a second-level procedure to combine subpaths into paths. We formalize subpath-based domination properties to establish the finite convergence and exactness of the column generation algorithm. We prove that the methodology can handle modeling extensions with heterogeneous charging costs (via dynamic re-optimization of charging decisions) and algorithm extensions to tighten the relaxation using ng-routes and limited-memory subset-row inequalities (via augmented domination criteria). Computational results show that the methodology scales to large instances, outperforming state-of-the-art column generation algorithms. From a practical standpoint, the methodology achieves significant cost reductions by jointly optimizing routing-scheduling and charging decisions and by capturing heterogeneous charging costs.

ME36

Summit - 428

Planning and Operations for Shared Mobility

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Kai Wang, Tsinghua University, Beijing, 100084

Co-Chair: Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States

1 - From Lab to Real World : Understanding the Practical Complexities of Decision Making for On Demand Transit Systems

Abhishek Dubey, Vanderbilt University, Nashville, TN, United States, Sophie Pavia, Aron Laszka, Ayan Mukhopadhyay, Samitha Samaranayake

While on-demand ride-sharing services have become popular in recent years, traditional on-demand transit services cannot be used by everyone, e.g., wheelchair-bound people. Paratransit services, operated by public transit agencies, are a critical infrastructure that offers door-to-door transportation assistance for individuals who face challenges in using standard transit routes. However, with declining ridership and mounting financial pressure, public transit agencies in the USA struggle to operate existing services. We collaborate with a public transit agency from the southern USA, highlight the specific nuances of paratransit optimization, and present a vehicle routing problem formulation for optimizing paratransit. We validate our approach using real-world data from the transit agency, present results from an actual pilot deployment of the proposed approach in the city, and show how the proposed approach comprehensively outperforms existing approaches used by the transit agency. We present a customizable, open-source algorithmic approach for paratransit optimization that allows for hard domain-specific constraints to be captured.

2 - Dynamic capacity planning for demand-responsive multimodal transit

Bernardo Martin-Iradi, ETH Zurich, Zürich, Switzerland, Kayla Cummings, Alexandre Jacquillat, Alexandria Schmid

Microtransit offers opportunities to enhance urban mobility by combining the reliability of public transit and the flexibility of ride-sharing. This paper optimizes the design and operations of a deviated fixed-route microtransit system that relies on reference lines but is allowed to deviate in response to passenger demand. We formulate a Microtransit Network Design (MiND) model via two-stage stochastic optimization. The model features a tight second-stage formulation thanks to a subpath-based representation of microtransit operations in a load-expanded network, which optimizes on-demand deviations between checkpoint stops. We develop a double-decomposition algorithm combining Benders decomposition and subpath-based column generation armed with a tailored label-setting algorithm. Using real-world data from

Manhattan, results suggest that our method scales to large practical instances, with up to 10-100 candidate lines and hundreds of stations. Comparisons with transit and ride-sharing benchmarks suggest that microtransit can provide win-win outcomes toward efficient mobility (high demand coverage, low operating costs, high level of service), equitable mobility (broad geographic reach) and sustainable mobility (limited environmental footprint). We provide an open-source implementation in an online repository to enable replication.

3 - Optimizing Relay Operations Toward Sustainable Logistics

Alexandria Schmid, MIT Operations Research Center, Cambridge, MA, United States, Alexandre Jacquillat, Kai Wang

Relay logistics decompose long-haul shipments into short segments through a pit-stop network, traveled by separate drivers. This paper develops optimization models, scalable algorithms and actionable insights to support and assess relay operations. A theoretical model highlights that relays can jointly improve drivers' working conditions, delivery speeds and distances traveled by smoothing out demand variability along high-volume pit-stop corridors. Next, we formulate a mixed-integer optimization model optimizes and coordinates the routing of orders, drivers and vehicles in large-scale relay logistics networks. To solve it, a multi-arc generation algorithm iteratively adds arcs to time-space networks by seeking a path of negative reduced cost within a relaxed and projected polyhedron. Extensive data-driven results show that (i) the algorithm outperforms state-of-the-art benchmarks and scales to large and practical instances; (ii) the optimization methodology can yield significant gains in relay logistics; and (iii) relay can provide benefits as compared to point-to-point operations, consistently with our theoretical findings. With our optimization methodology, relay operations can ultimately provide win-win-win outcomes in logistics toward the ``3 E's'' of sustainability: equity (working conditions), economy (delivery lead times, operating costs), and environmental footprint.

4 - Value of Sharing in Robots-as-a-service Operations

Kai Wang, Tsinghua University, Beijing, China, People's Republic of

``Robots-as-a-Service'' (RaaS) is an emerging operating model that promotes automation adoption through sharing. For example, FarmWise, a leader in robotics for agriculture, operates a fleet of weeding robots that serve many geographically dispersed users. Such RaaS platforms allocate robotic assets to users (a routing-scheduling problem) and assign tasks to these robots (a workload planning problem). This paper develops optimization models, practical insights, and scalable algorithms to support these decisions. First, a theoretical model analyzes the interplay between routing optimization and workload balancing. It emphasizes the two-sided value of sharing, i.e., the value of sharing robots across users, sharing workload across robots, and the necessity to combine both to achieve the best of both worlds. Next, this paper presents a mixed-integer optimization model to facilitate RaaS operations in practice. To solve this model efficiently, we develop an ``explore-and-bound'' algorithm that iterates between a restricted master problem and a convexity-based relaxation, while iteratively adding ``exploration cuts'' to visit new solutions. Computational results demonstrate that this methodology yields higher-quality solutions than several benchmarks and can scale to large and practical instances. By using data from FarmWise, we show that the proposed optimization approach can yield significant benefits, including a reduction in travel times (with environmental benefits) and overtime work (with social benefits), enabling a higher utilization of the robotic fleet.

ME37

Summit - 429

Mathematical Innovations in Curb Space Management, Pricing, and Service Operations

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Jisoon Lim, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Neda Masoud, University of Michigan Ann Arbor

1 - Spatiotemporal Pricing of Curb Space for Improving Operator and User Utilities

Jisoon Lim, University of Michigan, Ann Arbor, MI, United States

Effective curb space management emerges as a critical consideration within the revolutionary urban transportation system. Facilitating various transportation operations is contingent on appropriate allocation and pricing strategies for available curb spaces. We propose a Stackelberg game-theoretical model to illustrate the dynamics of curb space operation accommodating diverse curb space usages, where the game leader operates the curb space by setting spatiotemporal pricing policy and space allocation for each curb space usage, and the multi-followers evaluate the policy by accepting or rejecting curb space participation. Applying spatiotemporal pricing allows the operator to adjust pricing strategies across different curb space zones in response to demand dynamics and allocation policies. We formalize this game as a bi-level nonlinear optimization problem and solve it using Lagrangian relaxation applied to a single-level reformulation model. Numerical experiments demonstrate the effectiveness of spatiotemporal pricing schemes in mutually benefitting curb space stakeholders and improving the game equilibrium. We further discuss integrating practical considerations to capture the intricacies of curb space user characteristics.

2 - The Effect of Dynamic AI Surge Pricing on Passengers’ Tipping Behavior in Ride-Hailing Service: An Empirical Evidence from Uber Platform

Xiaohui Liu, Department of Information Systems and Analytics, School of Computing, National University of Singapore, Singapore, Singapore, Yujing Ma, Liang Zhang, Wei Ma, Hock-Ha Teo

Tipping on the Uber platform not only exerts a considerable influence on the economic remuneration of drivers but also serves as a vital social lubricant to enhance drivers’ job motivation and satisfaction. Accompanying the Uber platform, the designed and adopted Artificial Intelligence (AI) surge pricing benefits the platform directly by increasing company revenue. However, its dynamics in adjusting trip fares also have the potential to diminish drivers' benefits by reducing the ratio of tips received from passengers. This study empirically investigates the externality of AI surge pricing on passengers' tipping behaviour on the Uber platform in New York City. Empowered by the two-stage least square (2SLS) econometric method, our empirical evidence demonstrates that the higher AI surge pricing adopted by the Uber platform will negatively affect the tipping ratio of passengers. A series of mechanism checks justify that passengers, as the payers for both dynamic AI surge pricing and voluntary tips, will tend to transform their algorithm aversion towards platforms into negative tipping behaviours towards drivers. Our findings imply that the company's profit-driven pricing strategy may not achieve a win-win situation by overlooking its indirect impact on drivers' rights. This study provides insights into labor management on ride-hailing platforms, particularly regarding safeguarding labor benefits amidst the prevalent adoption and use of AI surge pricing in ride-hailing services.

3 - Economic Analysis of on-Street Parking with Urban Delivery

Zhengtian Xu, The George Washington University, Washington, DC, United States, Xiaotong Sun

The surge in online shopping has dramatically increased the demand for short-term curb access for package pickups and deliveries, leading to heightened competition for limited curb space. This study addresses the problem of how the unique parking demand of deliverers, particularly their parking duration for delivery attempts linked to parking space availability, affects the dynamics of urban curb parking systems. We develop continuum models of a curb parking system and perform analytical analyses to understand the dynamics and steady-state properties of the system under the influence of increased urban deliveries. We conduct comparative statics to examine how various curb management measures, such as pricing, parking duration caps, and dedicated delivery bays, influence the equilibrium conditions, followed by comparisons of the theoretical capacity of these measures. The results suggest that pricing strategies for metered parking to general parkers prove to be more efficient and flexible compared to other interventions. Particularly, when curb parking pricing is optimally calibrated, the necessity for dedicated delivery bays diminishes. Our findings provide valuable insights into the distinct impact deliverers have on curb parking dynamics, highlighting the need for a strategic reevaluation of current management practices.

4 - Adaptive Parking Pricing using Graph Neural Networks

Fatemeh Sadeghi, York University, Toronto, ON, Canada

Like many cities, Toronto has limited parking availability and high demand, which makes searching for parking spaces cumbersome. Rather than spending considerable time in an almost fully occupied parking space near the desired destination, a less expensive but farther parking location within walking distance may be an alternative for drivers. To balance the demand and the supply of parking zones, this study proposes an adaptive pricing algorithm for on-street parking locations in close vicinity, i.e., increasing the parking cost in a crowded spot. An iterative optimization algorithm is introduced to determine the optimal price for individual parking locations. The number of occupied parking locations (NOPS) is estimated using the transaction data provided by the Toronto Parking Authority (TPA), which is financially efficient and easily accessible. The TPA data comprises arrival and departure times and dates during May, June, and July 2019 and 2022. The Graph Neural Network (GNN) is specifically trained by the TPA data to predict the NOPS after updating the price profile. Input features for each parking location are categorized into sociodemographic data from the City of Toronto and amenity data obtained from the Open Street Map. The GNN framework captures spatial relationships among parking locations. A linked-based algorithm is proposed for the adaptive pricing algorithms. It shows that, after 30 iterations of price increase, the number of unbalanced parking locations has decreased from 2189 to 2134 out of 2572 comparisons and the average occupancy rate has increased from 27% to 38%.

ME38

Summit - 430

Aircraft and Crew Planning

Invited Session

Aviation Applications

Chair: Pushpendra Singh, Dartmouth College, Hanover, NH, United States

1 - Anticipating uncertainty in flight planning and scheduling

M. Selim Akturk, Bilkent University, Ankara, Turkey

We introduce a new scheduling approach for the flight planning and scheduling problem. Our main focus is to anticipate possible disruptions of an aircraft due to its mechanical failures along with possible delays that might be caused by the congestions of a flight at the origin and destination airports. Consequently, we could minimize the high opportunity costs caused by the flight cancellations and passenger inconvenience caused by delays in the schedule. We show that by allowing a small deviation from the minimum cost schedule at the schedule generation phase, the potential recovery costs in the future can be reduced significantly.

2 - Assessment of maintenance slot policies for an aircraft fleet using Value of Information in a Condition-Based Maintenance environment

Iordanis Tseremoglou, Delft University of Technology-Faculty of Aerospace Engineering, The Hague, Netherlands, Dimitrios Zarouchas, Bruno F. Santos

Aircraft maintenance is pivotal for airline safety and profitability. Following advancements in sensor technology and prognostics, aircraft maintenance is transitioning from traditional preventive and corrective approaches to Condition-Based Maintenance (CBM). However, the accuracy and uncertainty of Remaining Useful Life (RUL) predictions present challenges for optimizing maintenance operations. Tailored slot policies are essential, with conservative policies for high-certainty predictions and more frequent slot policies for uncertain RUL predictions. In this context, optimizing CBM necessitates quantifying its benefits, asking "how much information is worth?" through Value of Information (VoI) analysis. This extends beyond accounting only for the component maintenance costs but also considering the costs of introducing extra maintenance slots. Integrating VoI into policy decisions allows airlines to evaluate trade-offs between sensor investments and slot practices. A novel approach is introduced to evaluate maintenance slot policies in a CBM environment, considering RUL predictions of varying accuracy and uncertainty levels. Historical airline data and the C-MAPSS dataset are used to simulate maintenance tasks and RUL predictions. A scheduling model using a Deep Reinforcement Learning algorithm calculates maintenance costs under traditional and prognostics-driven approaches, extracting VoI associated with integrating RUL predictions. Results show that VoI varies significantly based on accuracy, uncertainty, component costs, and slot policy. The conducted VoI analysis provides useful insights that can aid decision-makers in selecting appropriate slot policies aligned with RUL predictions accuracy and uncertainty. Overall, this research advances CBM optimization for aircraft fleets, offering practical insights for enhancing maintenance operations in the aviation industry.

3 - Tail Assignmet-Driven Aircraft Routing Problem

Rubén Jiménez Moreno, Rey Juan Carlos University, Fuenlabrada, Spain, Manuel Fuentes González, Luis Cadarso

Airline planning confronts complex optimization challenges due to limited data availability, addressing aircraft routing and tail assignment across various timeframes, from months to days before operations commence. Operational fluctuations like flight cancellations necessitate continuous updates to rotations and tail assignments, demanding a flexible approach for seamless planning-to-operation transitions.

This research seeks to merge aircraft routing and tail assignment problems to address flight cancellations and maintenance-tail conflicts while optimizing cost-effective flight rotations. We propose a mixed-integer linear programming model, utilizing a directed acyclic graph representation, with tasks as nodes and connections as arcs. To boost efficiency, we introduce a clustering method grouping tails with similar characteristics.

Our computational experiments focus on a two-day planning horizon within a European airline network. The results illustrate the model's capability to maintain routing feasibility, minimize rotation changes, and meet specific maintenance criteria. This integrated approach offers a robust solution for navigating the dynamic airline operational landscape, ensuring efficient resource utilization and operational continuity.

Print Close

ME39

Summit - 431

Improving Rail Yard Management

Invited Session

Railway Applications

Chair: Marc Meketon, Oliver Wyman, 1 University Square Dr, Princeton, NJ, 08540, United States

1 - An Operational Perspective on AI Driven Switching Software

Kevin Murphy, Cedar AI, Seattle, WA, United States

As organizations increasingly rely on complex networks, the need for efficient traffic management and resource allocation has never been more critical. AI-driven switching software leverages machine learning algorithms to optimize data flow, reduce capacity constraints, build fluidity within the yard, and enhance overall network performance. We will examine key operational benefits, including bowl compliance and the positive impact it makes on yard operations.

2 - Evaluating Railway Network Dynamics with Multi-yard Simulation and Mainline Interaction

Tyler Dick, University of Texas at Austin, Austin, TX, United States, Jiaxi Zhao

When considering the performance, capacity and resiliency of a freight railway network, much attention is given to modeling train delay on mainline links, with a secondary emphasis on modeling the yards and terminals that serve as network nodes. Mainline links are typically modeled separately from yards and terminals, making it difficult to gain a holistic perspective on network performance and the interactions between these disparate facilities. However, interactions between mainlines and yards have been observed to play a critical role in the propagation of congestion across the rail network during historical disruption events, and are thus key to understanding railway network resiliency. To meet this industry need, the objective of this research is to simulate and analyze the dynamic performance and interaction between multiple classification yards and connecting mainlines within a railway network. To accomplish this goal, multiple instances of a classification yard model and a simplified mainline capacity model were integrated into a single AnyLogic simulation. The resulting rail network simulation model is used to investigate and quantify rail network recovery patterns after disruptions to yard operations, providing insights into the factors influencing rail network resiliency and the operational consequences of a yard disruption within a broader network scope. The resulting model and findings can help railway practitioners better understand the impact of operational or infrastructure changes in one terminal on other terminals and connecting mainlines, and make optimal decisions regarding allocating resources, developing train plans, and implementing strategies to mitigate delay after incidents in classification yards.

3 - Modeling Rail Yard and Network Operations

Roger Baugher, TrAnalytics, LLC, Duluth, GA, United States

A computer simulation tool, AnyLogic, has been used to model both yard and network operations. The yard model enables a user to create or import a switching problem and specify a solution which can be compared with an optimal solution generated by Cedar AI's Optiswitch software. The network model is limited to a simple configuration of four yards, yet a variety of operating plans and yard performance parameters can be specified and their impact on trip times evaluated. The intent is to acknowledge the probabilistic nature of railroad operations, a different approach from many models and trip plan systems that treat trip time as deterministic.

4 - Intermodal Terminal Layout Optimization Considering Efficiency and Sustainability

Qianqian Tong, University of Texas at Austin, Austin, TX, United States, Tyler Dick

In 2023, railroads in the United States transported 12.5 million intermodal containers and trailers, accounting for nearly 52% of carloads of total traffic. As railroads move towards decarbonization goals, and traffic from environmentally-conscious shippers shifts from highways to railways, the capacity and investment required to ensure the sustainability and efficiency of intermodal shipments is of prime concern both on the mainline and at terminals. Addressing these issues for road-rail intermodal terminals at a strategic level requires a preliminary understanding of the optimal terminal layout and resource levels to facilitate transshipment of increasing container volumes while promoting performance, reliability, efficiency and environmental friendliness. To better understand the role of expanding intermodal terminals in a low-carbon and energy efficient transportation system, we first propose a mixed-integer programming model and algorithm to solve for optimal facility layout and components for a target throughput volume, and second develop an intermodal transshipment facility simulation model to quantify the systematic impact of various factors on intermodal terminal throughput capacity, performance and emissions. Applying our model, we can maximize intermodal terminal capacity within specified land area constraints while balancing handling efficiencies and carbon emissions. The results can inform strategic investment decisions regarding terminal equipment and facility expansions to promote the sustainability of the overall intermodal supply chain. Meanwhile, this two-stage method also offers insights into identifying and alleviating bottlenecks arising from the intermodal terminal layout and allocated resources at the planning and operations level.

ME40

Summit - 432

BOM Best Working Paper Competition

Award Session

Behavioral Operations Management

Chair: Ignacio Rios, The University of Texas at Dallas, Richardson, TX, 75080, United States

1 - BOM Best Working Paper Competition - Welcome

Ignacio Rios, The University of Texas at Dallas, Richardson, TX, United States

This is Welcome and introduction to the award.

2 - Enhancing Volunteer Retention: The Role of Experienced Volunteers

Vinit Tipnis, Kelley School of Business, Indiana University Bloomington, Bloomington, IN, United States, Christopher Chen, Fei Gao

Volunteers play a pivotal role in the daily operations of nonprofit organizations (NPOs), yet retaining volunteers remains a critical challenge for NPOs. This study explores a cost-effective approach to improving volunteer retention by leveraging the potential of the existing pool of experienced volunteers. We examine how interactions between experienced and new volunteers impact the latter's retention behavior. Using data of 45,315 volunteers at a food bank over a four-year period, we explore interaction patterns among the volunteers. Our findings indicate that the presence of experienced volunteers in a volunteering shift increases new volunteer retention by 15.5%, with a stronger 71% increase for group volunteers. However, we find that increased familiarity among experienced volunteers negates this positive effect. This negative effect is dependent on the uncertainty inherent in the tasks undertaken by the volunteers. Our study contributes to the volunteer management literature by highlighting the importance of social interactions for volunteer retention and how NPOs can utilize their existing experienced volunteers to enhance new volunteer retention.

3 - Green E-commerce: Environmental Impact of Fast Delivery

Chenshan Hu, Washington University in St. Louis, Saint Louis, MO, United States, Xiaoyang Long, Jiankun Sun, Dennis Zhang

It is well-established that faster delivery in e-commerce increases consumer demand. However, the impact of faster delivery on how consumers place orders (their order frequency and basket sizes) and the subsequent environmental implications are not known. To investigate these questions, we leverage a quasi-experiment involving the opening of a new local warehouse by Alibaba Group, which led to a half-day improvement in the delivery speed for local orders. Through a difference-in-differences analysis, we find that the delivery speed improvement not only increased consumers' monthly purchasing amount by 6.70%, but also increased monthly order frequency by a higher percentage (i.e., 7.74%) and reduced the average order basket size by 0.79%. These results collectively suggest that with faster delivery, consumers purchase more on the platform but do so in more frequent and smaller orders, which implies more packaging and transportation costs for each unit of product sold. Based on these results, we conduct a detailed calibration using both public and company-specific data to estimate the increase in the platform's carbon emissions due to faster delivery. We also explore and identify two mechanisms contributing to the phenomenon: order-splitting and category expansion. We combine these insights with heterogeneous treatment effect analysis to derive managerial implications for the e-commerce platform. In particular, we find that for a platform that implements a threshold shipping policy, raising the free shipping threshold may be more effective than raising the shipping fee to reduce the environmental and operational costs associated with faster delivery.

4 - Delegated Innovation Scouting when Success is Rare: A Behavioral Investigation

Jochen Schlapp, Frankfurt School of Finance & Management gGmbH, Frankfurt Am Main, Germany, Mirko Kremer

Firms involved in the widespread practice of (delegated) innovation scouting face a crucial challenge: How to motivate innovation scouts to engage in extensive search for innovative projects given that search success oftentimes eludes them? Motivated by common scouting practice, we build a principal-agent model to study the optimal structure of formal incentive schemes as well as the performance of contract schemes that rely on discretionary and non-contractible rewards. We then test our key predictions in a controlled laboratory experiment that varies scouts' success chances and their contract type. We find that, irrespective of contract type, performance distortions are strongest in low success probability environments, due to poor incentive design by managers and poor search decisions by scouts. Our results are suggestive of two behavioral biases that reinforce one another when success is rare: (i) superficial fairness, which prevents managers from offer- ing the steep incentives needed to motivate scouts; and (ii) biased probability judgments, which lead both managers and scouts to overweight their success chances. Based on our findings, we derive actionable rec- ommendations on how to improve the design of innovation scouting processes in practice.

5 - Modeling the Influence of Unpaid Effort in Delivery Agents' Choices in Last-Mile Platforms

Lina Wang, The Pennsylvania State University, State College, PA, United States, Stanley Lim, Elliot Rabinovich

We examine decision-making dynamics in the choices that delivery agents make among jobs available in last-mile platforms. Jobs at these platforms comprise the delivery of multiple orders along routes and their compensation typically accounts for the number of deliveries and the length of the routes involved in making these deliveries (from the first to the last delivery). Rarely, do jobs compensate agents for the distance they must travel to make the first delivery in the routes. Traveling this distance (typically referred to as stem distance, deadhead, or empty miles) often constitutes a significant portion of the effort agents put in to complete these jobs. We seek to determine how this unpaid effort influences job selection by agents relative to other job attributes. We obtained job selection data from a last-mile delivery platform and, based on this data, used a structural framework to estimate a dynamic choice model that accounts for the effects of stem distances and other job attributes on agents' job selection decisions. This model also considers the uncertainty agents face regarding the availability and quality of future jobs by accounting for the option value of waiting as well as dynamic selections among agents. We find that an increase in stem distances markedly reduces the likelihood of job selection, overshadowing job attributes such as payout amounts and delivery lead times.

ME41

Summit - 433

Advancements in Graph Learning and Stochastic Process Optimization

Contributed Session

Chair: Yoonsik Jung, Korea University, Seongbuk-gu, Korea, Republic of

1 - Representation Learning for Event Sequences from a Diverse Range of Stochastic Processes with Attention-based Neural Networks

Francisca Quijada, Arizona State University, Tempe, AZ, United States, Micaela Mercado, Robert Wickham, Kristin Ferguson, George Runger

Event sequences are found in multiple domains where systems generate transactions or encounters over time. Their study is relevant to identify patterns and understand system behavior. Insights about future actions or common patterns benefit from an analysis.

Event sequences can be addressed via feature engineering (FE) and traditional predictive models. FE can include temporal and non-temporal attributes to represent sequences, but FE is a heavy and complex process requiring in-depth domain knowledge and may not fully represent the underlying temporal dependencies between events. We present a methodology for event sequences that vary in length and where events may not occur at periodic intervals. The methodology captures temporal and entity-level static features and compares to traditional FE for several learning tasks. The problem is modeled to aid policy decision-makers and practitioners in the homeless services system and is among the first modern machine-learning applications for event sequence prediction in this domain.

2 - Graphlet Proximity-based Negative Sampling for Graph Contrastive Learning

Myungsoo Choi, Ajou University, Suwon, Korea, Republic of, Hyunjung Shin

Given a set of graphs, labeling each graph is a challenging task: a label indicates whether a protein has a specific function, a chemical molecule is aromatic, and so on. Recent work has aimed to address this challenge by learning unsupervised representations through graph contrastive learning, which places positive pairs close together and negative pairs far apart in the embedding space. However, finding negative graphs that are not false negatives remains a critical difficulty. To tackle this issue, we propose Graphlet Proximity based Negative Sampling, which performs screening on the entire set of graphs using the kernel matrix of graphlets to avoid false negatives. Experiments show that the proposed method can significantly improve the performance of graph contrastive learning by efficiently filtering out false negatives.

3 - Enhancing Graph Contrastive Learning with Edge Distribution and Node Mutual Information

Jeongheun Yeon, Ajou University, Suwon, Korea, Republic of, Hyunjung Shin

Labeling nodes in a graph is time-consuming and expensive. Graph contrastive learning, an emerging method within self-supervised learning, facilitates graph representation learning without relying on label information. Since the labels are sparse or a few, the graph embedding is important for contrastive learning. Given a node of interest, called an anchor, the embedding is determined by selecting the positive samples that are likely to belong to the same class and the negative samples that are unlikely to belong to the same class.

In this study, we propose a novel graph contrastive learning utilizing edge distribution and node mutual information. The former is used to screen out the samples that are likely to be 'true' positive samples or 'true' negative samples. The latter is then used to select the samples that are likely to be near the class-boundaries. Embeddings become more robust with the former and more sophisticated with the latter, which further enhances contrastive learning. Experiments show that the proposed method outperforms the comparison methods by taking advantage of the more robust and sophisticated embedding results.

4 - On a Generalization of the Information Collection Problem

Nastaran Oladzad-Abbasabady, University at Buffalo-SUNY, Amherst, NY, United States, Esther Jose, John Becker, Rajan Batta

We study an Information Collection Problem (ICP) on a graph, where nodes hold valuable information. An agent, with adjustable speed and a given effective range, can collect this information by passing close enough to a node. Nodes are guarded and hence pose a threat to the agent. The objective is to find the path along which the maximum information can be collected, before the first detection occurs. A three-phase approach is presented to tackle this problem. First, we preprocess the graph to identify collectable information along different edges. Second, we develop a Mixed-Integer Non-Linear Programming (MINLP) model to minimize the risk of detection for the agent along each edge. The MINLP assumes that the agent can complete the mission before detection. To enhance the MINLP solution, a heuristic method is introduced in the final step. This method aims to maximize the "expected" information collected by considering agent's detection during the mission.

5 - Overcoming Spatial Feature Limitations in Image Classification through Convolutional Logical Analysis of Data

Yoonsik Jung, Korea University, Seoul, Korea, Republic of, Hong RYOO

In the realm of image classification, Logical Analysis of Data (LAD) utilizes RGB values of pixels as features, leveraging spatial information to classify images. While this approach demonstrates high performance in images with consistent patterns, it suffers from weakened evidence for each pattern when factors for classification are dispersed across various image locations. This drawback is not exclusive to LAD but extends to traditional machine learning methodologies and even neural networks like CNNs.

To address the limitations of spatial features, we aim to classify images based on how well the kernel's shape is embedded within the image data, utilizing convolution operations and optimization models. Toward the end, we propose optimization-based data analytics and LAD pattern generation and report promising results from preliminary experiments.

ME42

Summit - 434

Simulation Methods and Applications

Contributed Session

Chair: Yuming Zhao, NEU, Shenyang, N/A, China, People's Republic of

1 - A Learning-based Branch-and-price Approach for Solving the Cutting Stock Problem in Steel Industry

Yuming Zhao, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, 110819, Shenyang, China, People's Republic of, Ying Meng, Lixin Tang

In this work we address a cutting stock problem inspired by real-world steel industry plate cutting process in aiming to improve production management and efficiency. A formulation of the model with two-dimensional guillotine cutting is introduced. We propose an algorithm based on branch-and-price method and a deep learning-based approach is applied to improve the algorithm. Experimental results of generated problem instances indicate the algorithm’s efficiency in practical application.

2 - Multi-Manned Mixed-Model Assembly Line Balancing: Integration of Parallel Workstations and Variable Task-Resource Requirements

Wencang Bao, Georgia Institute of Technology, Smyrna, GA, United States, Leon McGinnis, Benoit Montreuil

The assembly line balancing (ALB) problem has been widely investigated for decades. In this paper, we focus on the integration of two popular extensions of ALB that can effectively improve resource utilization rates: parallel workstations (PW), and variable task-resource requirements (VTR). In our context, VTR includes both the choices of resource types and the decision on the resource quantity. In this research, we integrate PW and VTR for mixed-model and multi-manned assembly lines where multiple workers are allowed to work concurrently in one workstation. We formulate the problem by proposing a mixed-integer optimization model to analyze their interactions. Additionally, we develop four separate models, ALB with PW only, ALB with VTR only, ALB with PW first then VTR, and ALB with VTR then PW. We carry out an experiment based on a real-world case study and compare the performance of the integrated model and separate models.

3 - Reduction of Production TIME Variability in the Flexible Work Areas of Mixed-Model Production Systems Using Efficient Particle Swarm Optimization

Hyeongbae Yoon, Ajou University, Suwon-si, Korea, Republic of, Jeonghan Ko

This research investigates a new approach for production system flexibility and related optimization algorithm. In high product variety, the same type of tasks for different product models has large variation in process times. Performing such a task type in the same workstation has limitations in reducing production time variability. To overcome this problem, this research proposes a novel concept of flexible areas in a production line. A flexible area consists of workstations that can perform different tasks depending on product models. To find solutions minimizing production time variability, an efficient particle swarm optimization (PSO) method is developed. In this PSO algorithm, particle properties are affected by product models, and particle states are updated selectively. The results show that the suggested flexible area approach can reduce production time variability effectively. The results also demonstrate the efficiency of the new PSO algorithm, compared to the standard PSO methods. The solution quality and computation times were within reasonable ranges although the number of product models increases. The proposed flexible area approach and efficient particle swarm optimization can contribute to the realization of more flexible production systems and productivity improvement.

4 - Bayesian Optimization with Active Constraint Learning for Accelerating Material Synthesis

GUOYAN LI, Northeastern University, Boston, MA, United States, Yujia Wang, Swastik Kar, Xiaoning Jin

In the realm of manufacturing, the optimization of process parameters is crucial to achieving the objectives of enhanced productivity and quality. The advent of advanced manufacturing technologies introduces complex process dynamics, wherein the relationship between process parameters and quality is often nonlinear and nonstationary. Identifying the optimal process parameters with minimal experimental effort is particularly challenging, especially when manufacturing costs limit the number of feasible experimental trials. Processes such as material synthesis and additive manufacturing are two examples of such complexities. Identifying the optimal process parameters with minimal experimental effort is particularly challenging, especially when manufacturing costs limit the number of feasible experimental trials. Therefore, it's desirable to develop a fast yet accurate method to identify optimal process parameters efficiently using the fewest number of trials.

This study addresses the complex challenge of identifying process parameters for optimal manufacturing outcomes in advanced manufacturing, where nonlinear and costly process-to-quality relationships prevail. We introduce a novel experimental design framework that energizes the optimization of process parameters and feasibility constraint learning with minimal human intervention. Our approach is grounded in two primary methodologies: (1) active multi-criteria sample for constraint estimation and (2) Bayesian optimization-based sample for optimal parameter identification. This integration facilitates the efficient discovery of globally optimal parameter settings and outperforms multiple benchmark models in constraint estimation accuracy. The framework's efficacy is demonstrated through application on a real-world case study involving the synthesis of 2D materials, demonstrating its potential to enhance manufacturing efficiency and quality in complex manufacturing processes significantly.

5 - Revolutionizing Simulation Paradigms with Generative Agent

Young-Jun Son, Purdue University, West Lafayette, IN, United States, Md Tariqul Islam, Seung ho woo, Kamelia Sepanloo, Seonho Woo

The continuous evolution of simulation methodologies has pivotal implications for manufacturing, supply chain management, healthcare, social sciences, and more. Traditional modeling paradigms, including discrete event simulation, system dynamics, agent-based, and physics-based modeling, have long been foundational tools for analyzing complex systems and processes. However, the advent of generative agents rooted in advanced artificial intelligence (AI) poses a transformative potential that could augment or even replace much of these conventional methods. There has been some transformative research in this field, primarily trying to bridge the gap and pave the way towards adopting generative agents for different applications. This presentation explores the developments in generative agents, emphasizing their potential to enhance diversity, adaptability, scale, and scope in simulations. Through a comprehensive review, we will explore the integration of generative agents within industrial applications, using case studies to illustrate their practical benefits and limitations. The discussion will also address the theoretical implications and potential changes in methodology design and implementation for simulation applications. This analysis seeks to provide insights into the future direction of simulation modeling, offering a deeper understanding of how generative agents can redefine the boundaries of traditional modeling approaches.

ME43

Summit - 435

Managing Organizations to Boost Productivity

Contributed Session

Chair: Changyu Men, KU Leuven, Leuven, Belgium

1 - Local knowledge and operational performance: Evidence from a digitized setting

Changyu Men, KU Leuven, Leuven, Belgium, Marijn Verschelde, Maud Van den Broeke

In digital control rooms, controllers' experience significantly impacts their performance. They accumulate expertise through repetition, termed local knowledge, given the stable and repetitive nature of their tasks. Alternatively, they broaden their skills by working across various workstations, termed neighbor knowledge. Our study investigates the role and impact of these types of knowledge on operational performance, and examines the portability of neighbor knowledge within digital control room settings. For this purpose, leveraging a purposefully built comprehensive dataset from a major European rail system operator, we analyze how workstation-based and train-based experiences influence controllers' operational performance, measured by train delay. Our findings reveal that accumulated work at the same workstation improves operational performance, underscoring the importance of local knowledge. Further, when focusing on operational performance regarding passenger trains, we observe higher performance for controllers with more experience in passenger trains, but also lower performance for controllers with more experience in freight trains, indicating potential drawbacks of neighbor knowledge on operational performance.

2 - Reliable Inference from Human-Centred Datasets with Accumulated Local Effects

Chitu Okoli, SKEMA Business School - Université Côte d'Azur, Paris, France

Despite the increasing promise of big data analysis, operations research and management science (ORMS) problems often involve important human-centred topics with relatively small datasets. This study aims to improve the interpretability and performance of statistical inference techniques used in such cases. We focus on the General Linear Model (GLM) and Generalized Additive Models (GAMs), which are often more reliable than machine learning techniques on smaller datasets. GLM techniques are highly interpretable, and GAMs often provide superior performance as they model nonlinear relationships without prior specification. However, both are prone to issues (collinearity in GLM and concurvity in GAM) that confound the interpretation of correlated predictors.

To address these concerns, we extend accumulated local effects (ALE), a popular technique for visualizing relationships in machine learning models. Our approach combines bootstrapped ALE estimates with new ALE-based effect sizes to produce confidence regions for the predictor variables whose effects can be reliably inferred beyond the sample. ALE provides enhanced interpretability to GLMs and GAMs and is immune to the confounding effects of correlated predictors. Thus, the best-performing model specifications can be readily interpreted without resorting to the artificial exclusion of correlated yet relevant variables. Our results demonstrate that with ALE-based inference, researchers can inductively search for relationships in data without *a priori* hypotheses while effectively ruling out spurious patterns that plague classic hypothesis testing. Whereas such exploratory research is common with machine learning on large datasets, our extensions to ALE bring such benefits to smaller datasets for many important ORMS challenges.

3 - A Generic Approach for Managing Uncertainty in Execution of Large and Complex Projects: Generic Agile Project Management (Gapm)

Homayoun Khamooshi, George Washington University, Washington, DC, United States

Over the past few decades, the concepts of risk and uncertainty rightfully have been the focus of attention for researchers and practitioners of project management alike. It has been even argued that a no risk project is not worth pursuing. While plenty of high-quality research has been generated little practical solutions have emerged. In this paper it is argued that there is no deterministic project management case; but there exists a spectrum of cases stretching from very unlikely deterministic (CPM)projects to very probabilistic scenarios of high-risk projects. The more probabilistic cases to the right of the spectrum require a very dynamic project management approach like Agile. The question to be answered is how we could deal with higher levels of uncertainty which potentially translate into risks in the execution of our projects. Could an agile methodology be prescribed for riskier projects? If yes, how does that approach impacts execution of projects? The paper provides a generic approach for addressing and managing uncertainty in execution of large and complex projects labelled as GAPM: Generic Agile Project Management.

4 - Deep Learning-based White-box Adversarial Attack using Multivariate Time Series Classification Model

Juheon Kwak, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

Time Series Classification is widely used in various fields, such as medical, manufacturing, and etc.. Recently, multivariate time series are usually used than univariate time series in classification and various models are used such as Transformer, Recurrent Neural Network(RNN), and etc.. Adversarial attack to time series classification is mainly black-box attack, which is an attack without using the structure and gradient of model. Substitute model is necessary to mimic original model for black-box attack. To make substitute model is so difficult, so black-box attacks are very challenging and uncomplete. However, white-box attack uses all of model structure, so the attack is very easy and fast to success. White-box attacks are usually used to multivariate time series regression. We proposed the adversarial attack method, that is a white-box attack method to multivariate time series classification model using white-box attack. 5 public datasets were used, and the Convolutional Neural Network(CNN), RNN, and Transformer were used as models. Adversarial examples using our method are very similar to original multivariate time series than existing adversarial attack methods. And the trained models are well fooled by our method than existing attack methods.

5 - A Distributionally Robust Approach for Managing Hazmat Emergency Response

Ginger Ke, Memorial University of Newfoundland, St. John's, NL, Canada, Jiahong Zhao, Jie Wu, Longfei Zhang

The potential tremendous threats posed by incidences associated with hazardous materials (hazmat) to the surrounding people and the environment necessitate a well-managed emergency logistics network, especially for random emergencies. Focusing on the distributional ambiguity of the severity level and link disruption, we herein propose a distributionally robust optimization (DRO) model to construct a robust and reliable hazmat emergency response system. Specifically, a joint probability underlying the emergency demand and link disruption is defined to reflect the different characteristics of these two types of randomness. Then, based on the deterministic case, the DRO counterpart is formulated to minimize the total system cost and the worst-case risk over uncertainties. For the optimal solution, the DRO model is reformulated and solved by a Benders decomposition approach. We apply this model and algorithm to a real-world case study and derive managerial insights from numerical experiments to improve practical hazmat emergency management.

ME44

Summit - 436

Derivative-free Simulation Optimization

Invited Session

Simulation Society

Chair: Sara Shashaani, North Carolina State University, Raleigh, NC, United States

Co-Chair: Hyungkhee Eun, North Carolina State University, Raleigh, United States

1 - How to address derivative calculation in stochastic bilevel optimization

Tommaso Giovannelli, Lehigh University, Bethlehem, PA, United States, Griffin Kent, Luis Nunes Vicente

Bilevel stochastic optimization formulations have become instrumental in a number of machine learning contexts such as continual learning, neural architecture search, adversarial learning, and hyperparameter tuning. Bilevel first-order methods require second-order derivatives that become impractical to compute in optimization or learning problems with a high number of variables or constraints.

In this work, we introduce a bilevel stochastic gradient method for bilevel problems with nonlinear and possibly nonconvex lower-level constraints. We also present a comprehensive convergence theory that addresses both the lower-level unconstrained and constrained cases and covers all inexact calculations of the adjoint gradient (also called hypergradient), such as the inexact solution of the lower-level problem, inexact computation of the adjoint formula (due to the inexact solution of the adjoint equation or use of a truncated Neumann series), and noisy estimates of the gradients, Hessians, and Jacobians involved. To promote the use of bilevel optimization in large-scale learning, we have developed new low-rank practical bilevel stochastic gradient methods (BSG-N-FD and BSG-1) that do not require second-order derivatives and, in the lower-level unconstrained case, dismiss any matrix-vector products.

2 - Efficiency Analysis of Simulation Optimization with Dynamic Stratification of Large Input Spaces

Pranav Jain, North Carolina State University, Raleigh, NC, United States, Sara Shashaani

Stratification is a powerful technique employed to reduce variance, particularly by capitalizing on the local dependence observed in simulation inputs and outputs. In our approach, we integrate stratification into an adaptive sampling procedure within a derivative-free trust-region-based optimization algorithm. This integration allows for the dynamic adjustment of stratification within the optimization framework, responding effectively to the evolving trajectory of the optimization process. Through rigorous analysis and experimentation, we show that our trust-region method, enhanced with stratified adaptive sampling, not only ensures almost sure convergence but also offers significantly improved work complexity when compared to its non-stratified counterpart.

3 - Calibrating Digital Twin via Bayesian Optimization: A root-finding strategy

Yongseok Jeon, North Carolina State University, Raleigh, NC, United States, Sara Shashaani

In this research, we study leveraging root-finding strategies combined with bayesian optimization, which is particularly useful in digital twin calibration tasks. Employing root-finding strategies requires modified acquisition functions and offers unique advantages over traditional minimization strategies. Often, root-finding problem can be reformulated into a minimization problem, and many practitioners have been using these benefits without fully exploring its implications. Such transformations have to be carefully handled, as they may lead to additional challenges, depending on the functional structure or the employed optimization algorithms. Our study suggests that this transformation is not generally recommended, especially when using a continuous surrogate model. We illustrate several motivating examples on this topic, followed by our empirical results through a range of calibration tasks.

4 - Comparative Analysis of Distance Metrics in Distributionally Robust Optimization for Queuing Systems: Wasserstein vs. Kingman

Hyungkhee Eun, North Carolina State University, Raleigh, NC, United States, Sara Shashaani, Russell Barton

This study examines the effectiveness of different metrics in constructing ambiguity sets for Distributionally Robust Optimization (DRO). Two main approaches for building ambiguity sets are the moment- and the discrepancy-based approaches. The latter is more widely adopted because it incorporates a broader range of distributional information beyond moments. Among discrepancy-based metrics, the Wasserstein distance is often preferred for its advantageous properties over ϕ-divergence. In this study, we propose a moment-based Kingman distance, an approximation of mean waiting time in G/G/1 queues, to determine the ambiguity set. We demonstrate that the Kingman distance provides a straightforward and efficient method for identifying worst-case scenarios for simple queue settings. In contrast, the Wasserstein distance requires exhaustive exploration of the entire ambiguity set to pinpoint the worst-case distributions. These findings suggest that the Kingman distance could offer a practical and effective alternative for DRO applications in some cases.

ME45

Summit - 437

Innovative Approaches to Healthcare Experimentation and Inference

Invited Session

Health Applications Society

Chair: Wanning Chen, University of Washington, Seattle, WA, United States

1 - Estimating and Improving Individualized Treatment Rules with An Instrumental Variable

Bo Zhang, Fred Hutchinson Cancer Center, Seattle, WA, United States

Individualized treatment rules (ITRs) are considered a promising recipe to deliver better policy interventions. One key ingredient in optimal ITR estimation problems is to estimate the average treatment effect conditional on a subject's covariate information, which is often challenging in observational studies due to the universal concern of unmeasured confounding. Instrumental variables (IVs) are widely used tools to infer the treatment effect when there is unmeasured confounding between the treatment and outcome. In this work, we propose a general framework of approaching the optimal ITR estimation problem when a valid IV is allowed to only partially identify the treatment effect. We introduce a novel notion of optimality called "IV-optimality". A treatment rule is said to be IV-optimal if it minimizes the maximum risk with respect to the putative IV and the set of IV identification assumptions. We derive a bound on the risk of an IV-optimal rule that illuminates when an IV-optimal rule has favorable generalization performance. We propose a classification-based statistical learning method that estimates such an IV-optimal rule, design computationally efficient algorithms, and prove theoretical guarantees. We then discuss extension of an IV-based approach to longitudinal settings. We apply our method to study which mothers would benefit from traveling to deliver their premature babies at hospitals with high level neonatal intensive care units.

2 - Randomization-Based Confidence Intervals for the Local Average Treatment Effect

Haoge Chang, Microsoft Research, Cambridge, MA, United States

We consider the problem of generating confidence intervals in randomized experiments with noncompliance. We show that a refinement of a randomization-based procedure proposed by Imbens and Rosenbaum (2005) has desirable properties. Namely, we show that using a studentized Anderson--Rubin-type statistic as a test statistic yields confidence intervals that are finite-sample exact under treatment effect homogeneity, and remain asymptotically valid for the Local Average Treatment Effect when the treatment effect is heterogeneous. We provide a uniform analysis of this procedure.

3 - Dyadic Reinforcement Learning

Shuangning Li, University of Chicago, Chicago, IL, United States

Mobile health aims to enhance health outcomes by delivering interventions to individuals as they go about their daily life. The involvement of care partners and social support networks often proves crucial in helping individuals managing burdensome medical conditions. This presents opportunities in mobile health to design interventions that target the dyadic relationship---the relationship between a target person and their care partner---with the aim of enhancing social support. In this paper, we develop dyadic RL, an online reinforcement learning algorithm designed to personalize intervention delivery based on contextual factors and past responses of a target person and their care partner. Here, multiple sets of interventions impact the dyad across multiple time intervals. The developed dyadic RL is Bayesian and hierarchical. We formally introduce the problem setup, develop dyadic RL and establish a regret bound. We demonstrate dyadic RL's empirical performance through simulation studies on both toy scenarios and on a realistic test bed constructed from data collected in a mobile health study.

4 - Online Learning in Matching Market via Matrix Completion

Zhiyuan Tang, The University of Texas at Dallas, Richardson, TX, United States, Wanning Chen, Kan Xu

Two-sided marketplaces have become increasingly important for facilitating interactions among customers with matching needs across various domains, such as online labor markets, dating apps, and ride-hailing platforms. To achieve specific matching goals, such as optimal or stable matching, platforms need information on pairwise matching outcomes, which can be formalized into a reward matrix. In large-scale markets, missing information in the reward matrix can pose significant challenges for platforms aiming to achieve these matching goals. In this work, we use matrix completion method to learn missing entries from incomplete reward matrix and expedite online learning in two-sided marketplaces. We provide the first minimax optimal Frobenius norm error bound of matrix completion with matching dependence structure and first infinity norm bound of matrix completion with matching dependence structure by a second stage enhancement procedure. By incorporating our method into bandit algorithm, we improve the regret rate in online learning settings.

ME46

Summit - 438

Sanjay and Panna Mehrotra Research Excellence Award Session

Award Session

Health Applications Society

Chair: Susan Lu, Purdue University, West Lafayette, IN, United States

1 - Managing Healthcare Information through Algorithms

Mehmet Ayvaci, The University of Texas at Dallas, Richardson, TX, United States

The successful delivery of medical care relies on effectively acquiring and managing information to support decisions. While algorithms can assist in this pursuit, their effectiveness is influenced by input data quality, surrounding processes, and user behavior. These dependencies create challenges in realizing algorithmic value, but also offer opportunities for researchers to improve care delivery. I describe my ongoing research and collaborations focused on using algorithms to address these challenges in healthcare settings.

2 - Improving Patient Access to Biopharmaceuticals

Tugce Martagan, Eindhoven University of Technology, Eindhoven, Netherlands

More than 8,000 biopharmaceuticals are in the global research and development pipeline to treat cancer, diabetes and many other diseases. These drugs are produced using biomanufacturing technologies. With increasing demand and competition, the industry is currently experiencing a growing need for cost-effective and robust production to deliver affordable medicines to patients.
Over the past decade, I have worked with several biopharmaceutical manufacturing companies, hospitals, and government agencies in Europe and the United States. In these projects, we have explored the application of artificial intelligence (AI) and operations research (OR) methods to improve efficiency, safety, and innovation. The real-world implementation of the developed AI/OR models significantly reduced costs and improved productivity with a 50% increase in batch yield and a 40% reduction in carbon footprint. In this presentation, I will reflect on these AI/OR implementations in the industry to advance human health and achieve sustainable development goals.

3 - Transplant health inequities research from an operations perspective

Mustafa Akan, Carnegie Mellon University, Pittsburgh, PA, United States

Health inequity permeates all corners of our health care ecosystem. However, there are many powerful tools to effect real change. In this talk, I review the recent research on deceased donor organ transplant that offers solutions to reduce inequity in access to transplantation and mitigate the impact of those inequities in society at large.

ME47

Summit - 439

Decision-analytic Modeling for the Effective Management of Healthcare Operations

Invited Session

Health Applications Society

Chair: Gizem Nemutlu, Brandeis University, Waltham, MA, United States

Co-Chair: Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

1 - Leveraging AI for SMASH: Social Media Analytics for Adverse Side Effect Hunting

Abraham Seidmann, Boston University, NEWTON, MA, MA, United States, Alon Bartal, Nava Pliskin, Kathleen Jagodnik

Adverse side effects (ASEs) caused by medications cost healthcare systems trillions of dollars annually. To tackle this issue, we have developed an analytical methodology called Social Media Analytics for Side Effect Hunting (SMASH), which employs AI to integrate social media analytics, Named Entity Recognition, and Graph Convolutional Networks. By analyzing social media, ChatGPT responses, and pharmaceutical data, SMASH is capable of uncovering potential ASEs that may not be documented in clinical trials. We have successfully applied SMASH to glucagon-like peptide one receptor agonists (GLP-1 RAs) medications used for treating diabetes and obesity, expected to reach a market value of $133.5 billion by 2030. Our 'Early Warning' SMASH methodology has discovered 21 potential ASEs overlooked during regulatory approval. Potential ASEs are validated by estimating their frequencies (F1-score of 0.79, AUC of 0.82). Our innovative work completely revolutionizes ASE discovery by leveraging cutting-edge AI-driven social media analytics. Our approach goes beyond traditional information systems research, enabling us to proactively mitigate risks for pharmaceutical companies, regulators, and healthcare organizations.

2 - Applications of a new hybrid network-compartmental simulation technique for epidemic modeling

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

We developed a new hybrid agent-based network and compartmental (MAC) simulation technique for epidemic modeling. Individuals in a population can either be represented as agents in a network or modeled in a compartmental model that follows a Markov processes structure. An evolving contact network algorithm, developed using deep neural network and optimization, maintains the contact dynamics between the two populations. We applied this simulation technique to multiple applications related to sexually transmitted diseases, including joint modeling interacting diseases that have widely varying disease burden, mechanistic modeling social determinants of health into disease predictions, and cluster detection methods for detection of new outbreaks.

3 - Appointment Scheduling of Outpatient Clinical Services under Uncertain Patient Flows

Yang Yang, Purdue University, West Lafayette, IN, United States, Frank Chen, Qi Feng, J Shanthikumar

We study the problem of appointment scheduling for outpatient clinical services. The patient arrival, appointment cancellation, and appointment attendance are all random, leading to highly variable patient flows in the service system. Analyzing the dynamic decisions is challenging because of the need for tracking the confirmed appointments within the scheduling window. For planning with a long horizon, we show that a probabilistic allocation policy can be efficiently computed by optimizing a closed-form function. The resulting long-term profit approaches the optimal profit asymptotically for scalable systems. For a limited planning horizon, we propose a sequential assignment process implemented with a pre-determined scheduling diagram. We then characterize conditions under which a scheduling diagram policy is optimal. An improvement from a probabilistic allocation policy or a same-day scheduling policy results in a scheduling diagram policy, and the scheduling diagram policy is optimal for two-day schedules. The computation of the proposed scheduling diagrams, through ranking the margins of scalar functions, is simple. An extensive simulation analysis suggests that the system can be efficiently managed through a combination of a two-day policy and an improved probabilistic policy with the former applied to a short scheduling window and the latter to a long scheduling window.

4 - The "Netflix" Model: A New Payment Model For Asymptomatic Disease Management

Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States, Zhaowei She, Yueran Zhuo, Jagpreet Chhatwal, Turgay Ayer

This paper presents a framework to analyze the effectiveness of Netflix-style contracts in incentivizing payers to increase proactive screenings and pharmaceutical companies to lower treatment prices for asymptomatic diseases such as hepatitis C, breast cancer, colon cancer, and ovarian cancer. These contracts allow payers unlimited access to treatments by making upfront fixed lump sum payments, effectively lowering the unit treatment price as more patients are identified and treated. The authors show that by offering Netflix-style contracts alongside traditional volume-based contracts, pharmaceutical companies can separate payers based on their screening ability, with high-ability payers choosing Netflix-style contracts and low-ability payers remaining in volume-based contracts. This incentive-compatible nonlinear pricing mechanism can encourage payers with high screening ability to increase proactive screenings and lower unit treatment prices for these payers. To support the Biden Administration's initiative to eliminate Hepatitis C, the authors conducted simulations using a micro-simulation model and empirical analyses employing the Almost Ideal Demand System (AIDS) model on Medicaid Drug Utilization data. The findings aim to contribute insights into optimizing screening and treatment strategies for Hepatitis C, aligning with national health objectives.

5 - Saving Lives: Unveiling the Cost-Effectiveness of Liver Cancer Surveillance in Alcohol-Related Cirrhosis Patients

Ali Hajjar, Prince Mohammed Bin Salman College of Business & Entrepreneurship, King Abdullah Economic City, Saudi Arabia, Jovan Julien, Gizem Nemutlu, Hayrullah Mert Sahinkoc, Mary Linton Peters, Fasiha Kanwal, Jagpreet Chhatwal

Alcohol-related cirrhosis is one of the main risk factors for liver cancer. Our objective was to evaluate the cost-effectiveness of biannual liver cancer surveillance in these patients. Employing a microsimulation model, we simulated the progression of liver cancer in individuals with alcohol-related cirrhosis. By integrating data on liver cancer incidence, tumor advancement, real-world adherence to surveillance, and associated economic and health-related quality of life factors, we compared the cost-effectiveness of biannual ultrasound and alpha-fetoprotein surveillance with no surveillance across the patient's lifespan. In alcohol-related cirrhosis patients who continue drinking despite cirrhosis, biannual surveillance remains cost-effective (with an ICER of $19,000/QALY). Similarly, for those who have ceased drinking, biannual surveillance proves cost-effective (with an ICER of $56,000/QALY). These results underscore the cost-effectiveness of biannual liver cancer surveillance in alcohol-related cirrhosis patients, regardless of their drinking status, emphasizing its importance in clinical practice.

ME48

Summit - 440

Supply Networks and Innovation Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Kedong Chen, Old Dominion University, Norfolk, VA, United States

1 - Firm-Level Small-Worldliness: Exploring the Structural Characteristics and Performance Effects

Yusoon Kim, Oregon State University, Corvallis, OR, United States, Hyunwoo Park

In social or inter-organizational network contexts, small-world structure has attracted much academic attention and is widely thought to enhance creativity or innovative productivity. That is, the so-called small-worldliness has been theorized and studied predominantly at the network levels, while we often observe the effects of such phenomenon at the individual actor levels, such as when we encounter otherwise very distant yet useful individuals or information. In this paper, we consider some individual-level structural properties that induce or increase chances for small-worldliness effects. While the extant literature suggests some network-level structural factors—i.e., cohesive clusters of local ties and nonlocal ties connecting those clusters—as main determinants of small-worldliness effects, at the actor level, we suggest different structural factors that correlate with increased innovative performance. For an empirical test of our thesis, we used network data from the pharmaceutical industry that shows small-worldliness properties and demonstrates much evidence of the effects. Further, we provide some implications of our findings for future research on industrial/supply networks and firm-level innovation.

2 - Industry Concentration and Supply Chain Reconstruction: An Empirical Study

Pankaj Kumar, Virginia Tech, Blacksburg, VA, United States, Xiaojin Liu, Kedong Chen, Feng Cheng

Amidst rising market uncertainties, firms recognize the increasing need to develop efficient, effective, and adaptable supply chains. We conduct an empirical investigation into how a focal firm's position in a concentrated product market affects its supply chain reconstruction, specifically in terms of vertical integration and supply-base differentiation.

ME49

Summit - 441

Data-Driven and Sustainable Mobility and Logistics System Design

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Sheng Liu, University of Toronto, Toronto, ON, Canada

1 - The Path to Green in Ride-Hailing

Soraya (Nadia) Fatehi, The University of Texas at Dallas, Richardson, TX, United States

Ride-hailing platforms are working toward achieving greener mobility. To this end, they need to implement policies that make driving zero-emission electric vehicles (EVs) an economically attractive choice for crowd drivers. In this study, we propose policies that help ride-hailing platforms meet their carbon-emission sustainability target. Optimal policies suggest that platforms should prioritize allocating demand to available EV crowd drivers, and if necessary, offer subsidies to gas drivers to encourage their participation during this transition to green where the availability of EV drivers is limited. A key finding in our study emphasizes the role of time-to-service guarantee that the platform offers to its riders. If the platform slightly relaxes its tight “on-demand” guarantee, the platform can significantly maximize its earnings while substantially reducing emissions, due to better utilization of EV drivers. Additionally, we compare different incentive structures currently employed by ride-hailing platforms to promote the adoption of EVs and show that, at optimality, they result in the same expected performance. Furthermore, our investigation into the role of EV-rental providers in the platform’s transition to green reveals that if crowd drivers cannot afford the upfront cost of owning an EV and have to rely on EV-rental providers, the strategy of high EV-supply utilization is not always optimal due to higher operating costs. This emphasizes the crucial role of governments in making EV-ownership an affordable option. To assess the practical effectiveness of the policies, we conducted a real-world simulation study in Tampa Bay, USA, using Uber travel times and Tampa Bay’s charging infrastructure.

2 - A Simultaneous Column-and-Row Generation Solution Method for Liner Shipping Network Design

Jun Xia, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhou Xu, Roberto Baldacci

The liner shipping network design (LSND) problem aims at creating a set of regular services or rotations of oceangoing ships for a carrier to transport containerized cargo among seaports. Existing literature has shown that finding an optimal solution for the LSND is particularly challenging, even for small-scale problems, because of the complex rotation structure and the joint decision-making on rotation design, fleet deployment, cargo routing, etc. This work presents a new set-partitioning-like formulation for the LSND problem with transshipment costs, featuring an exponential number of variables and constraints. The formulation considers decisions on general rotation configurations of various service components, including but not limited to the ship type, sailing speed, and service frequency. To address transshipment costs correctly, the proposed formulation is associated with exponentially many rotation-dependent variables and constraints, with the result that even solving the linear programming relaxation is a big challenge. To tackle such a challenge, we propose a new simultaneous column-and-row generation (SCRG) solution method and novel speed-up techniques. We embed the SCRG solution method within a branch-and-price algorithm to develop a new exact method for the LSND and test it on two main LSND variants based on different rotation configurations. Results from computational experiments validate the effectiveness and efficiency of the newly proposed solution method. Our work not only enriches the solution methods for the LSND but also enhances the SCRG solution method in general and extends its practical applications.

3 - Machine Learning-Augmented Optimization of Large Bilevel and Two-Stage Stochastic Programs: Application to Cycling Network Design

Bo Lin, University of Toronto, Toronto, ON, Canada, Timothy Chan, Shoshanna Saxe

Motivated by a cycling infrastructure planning application, we present a machine learning approach to solving bilevel programs with a large number of independent followers, which as a special case includes two-stage stochastic programming. We propose an optimization model that explicitly considers a sampled subset of followers and exploits a machine learning model to estimate the objective values of unsampled followers. Unlike existing approaches, we embed machine learning model training into the optimization problem, which allows us to employ follower features that cannot be represented using leader decisions. We prove bounds on the optimality gap of the generated leader decision as measured by the original objective that considers the full follower set. We develop follower sampling algorithms to tighten the bounds and a representation learning approach to learn follower features, which are used as inputs to our machine learning model. Through numerical studies, we show that our approach generates leader decisions of higher quality compared to baselines. Finally, we perform a real-world case study in Toronto, Canada, where we solve a cycling network design problem with over one million followers. Compared to the current practice, our approach improves a transportation metric by 19.2% and can lead to a potential cost saving of $18M.

4 - Staffing Planning for Last-Mile Delivery Drivers

Tamar Cohen-Hillel, UBC Sauder School of Business, Vancouver, BC, Canada, Tolga Cezik, Liron Yedidsion

Staffing planning for last-mile delivery drivers is a crucial process in the logistics industry. It involves determining the optimal number of drivers needed each week to handle the expected volume of deliveries within a specific time frame. In this problem, one must determine the headcount over a planning window, accounting for cross-time constraints, while facing volume uncertainty. We illustrate that this problem is NP-hard and propose a DP-based approximation algorithm. We further show that the approximation algorithm achieves a tight optimality gap. Finally, we show, through computational experiments, that the algorithm performs close to optimality.

5 - Bike Lane Network Expansion: Traffic, Emissions, and Equity Implications

Keji Wei, Tongji University, Shanghai, China, People's Republic of, Sheng Liu, Jingwei Zhang, Xiaoyun Niu

Bike lanes are expanding rapidly in cities worldwide and serve as a major infrastructure policy lever for municipalities to promote cycling and achieve net-zero goals. Meanwhile, the congestion and equity concerns over bike lane network expansion are mounting, and a systematic evaluation of the environmental implications of bike lanes is lacking. Based on real-world data, we develop a mathematical model and empirical framework to estimate the impact of bike lane expansion on urban traffic, carbon emissions, and how different population groups are affected.

Our results indicate significant carbon emission reductions from bike lanes: 50 miles of new bike lanes can decrease emissions by 3.69\% in downtown Chicago. This is a combined effect of demand shift and traffic reassignment. This results in rush hour savings amounting to millions of dollars per year. However, the bike lane expansion is not distributed evenly, with more newly added bike lanes allocated to higher-income communities. As a result, low-income communities face lower coverage, less carbon reductions, and more limited access to bike lane infrastructure than high-income communities. On the other side, lower-income communities still enjoy the significant benefit of improved travel speed due to the network effect, which shows the nuanced equity implications of bike lane network expansion.

ME50

Summit - 442

Empirical Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Mohamad Soltani, University of Alberta, Edmonton, AB, Canada

1 - The Impact of Heat Hazard on Couriers' Health

Susan Lu, Purdue University, West Lafayette, IN, United States

To be completed...

2 - How a Recalculation of Workload Alters the Perception of Workload in Nursing

RJ Niewoehner, Kelley School of Business, Indiana University, Bloomington, IN, United States, Harshita Kajaria Montag

Nurses play a pivotal role in patient care, but only if systems can hire enough of them and convince them to stay. Though nurses make up the largest segment of the healthcare workforce, we find ourselves in the middle of a nursing crisis, and a shortage of nurses inevitably leads to overworked nurses. As a proxy for workload, most health systems carefully monitor the “nurse-to-patient ratio." We propose a fundamentally different tactic: using a painstakingly detailed workload tool, we observe how managers react to more accurate definitions of workload, including how they can better allocate and balance work among their staff, all while measuring how the workforce perceives the adjustments in real time. We leverage data from the field and the archive surrounding the introduction of our workload tool. In the field, we surveyed the “perceived workload" of nurses via the NASA-TLX, an industry standard. From the archive, we pulled observations from 71 unique nursing units operated by a large US academic health system. Prior studies generally measure workload using the nurse-to-patient ratio, whereas we consider the individual tasks required for each nurse-to-patient interaction. Moreover, our study monitors how managers react to new workload information and how frontline employees react to changing workload conditions.

3 - First, Do No Harm: Do Staffing Shortages Drive Abuse and Malfeasance in U.S. Nursing Homes?

Jong Myeong Lim, Miami Herbert Business School, Coral Gables, FL, United States, Kenneth Moon, Minjie Park

The U.S. nursing home industry has suffered decades-long understaffing and high worker turnover. Not only does understaffing undermine the quality of nursing home services and care, we empirically examine the novel possibility that it may incentivize unethical practices that cause patients harm. We study the causal linkage between nursing home staffing and chemical restraints, i.e., the improper use of antipsychotic medications to sedate difficult patients. Using nationwide data obtained from the Centers for Medicare & Medicaid Services through Freedom of Information Act requests, we conduct causal analyses that exploit local nursing market wages and changes in local minimum-staffing regulations as exogenous sources of variation in nursing homes' staffing levels. We find that every additional 15 minutes of daily nurse staffing hour per resident reduces a nursing home's usage of antipsychotic drugs by 2.1 residents, on average. Our causal estimates provide supporting evidence for the new CMS nursing home minimum staffing standards finalized in April 2024. Based on our estimates, the new minimum staffing rule will reduce nursing homes' usage of antipsychotic drugs by 36,495 residents, or 16.5%, nationwide. Our findings demonstrate a robust and widespread causal relationship between staffing and the use of potentially improper and harmful practices in healthcare operations.

4 - Capacity Expansion in Emergency Department Network: Impact of Opening Hospital and Hospital-Based Emergency Department on Patient Demand Shift and Operational Performance

Eric Park, Wake Forest University, Winston Salem, NC, United States, Bob Batt, Hyun Seok Lee, Timothy Rainer

We study how patients respond to easier access to emergency care in the form of new EDs opening closer to their location. Our study exploits individual patient’s decision on ED visit choice and tendency and its impact on care outcomes and operational performances due to shift in overall demand and subsequent change in ED congestion levels in two different settings. One ED/hospital opening in a region without any public hospital within 25 km and another ED/hospital opening in a region with already two other EDs within 10 km.

ME51

Summit - 443

Control of Queueing Systems with Applications to Service and Production Operations

Invited Session

MSOM: Service Operations

Chair: Nasser Barjesteh, University of Toronto, Toronto, ON, Canada

Co-Chair: Amir Alwan, The University of Chicago Booth School of Business, Chicago, IL, United States

1 - A Unified Fluid Model for Large Service Systems with Patience-Dependent or Delay-Dependent Service TIMES

Deniz Simsek, Northwestern University, Evanston, IL, United States, Achal Bassamboo, Ohad Perry, Chenguang Wu

We consider a single-pool many-server queueing system, assuming the service time of each customer depends on the delay of that customer in queue. Such dependence can be due to the customers having finite patience for waiting that depends on their individual service requirement, or due to having their service-time distribution be a function of the time spent in queue. We refer to the former dependence mechanism as “exogenous dependence” and to that latter as “endogenous dependence.” Since exact analysis of the stochastic system under either dependence mechanism is intractable, we propose a deterministic approximation for the (mean) queueing dynamics, referred to as a Unified Fluid Model (UFM). For constant arrival rates, we characterize conditions for the existence of a unique stationary point for the UFM and prove that those conditions always hold when the dependence is exogenous. However, the UFM may possess multiple equilibria, with each equilibrium point being either locally stable or unstable. The implication of the multi-stability of the UFM for the stochastic system is two-fold. First, the stochastic fluctuations in steady state may be an order of magnitude larger than the typical fluctuations in many-server queueing systems. Second, an underloaded or critically loaded system may experience congestion collapse, namely, it becomes severely congested.

2 - Delay Information Sharing in Two-Sided Queues

Mustafa Akan, Carnegie Mellon University, Pittsburgh, PA, United States, Mohammad Delasay, Siddharth Singh

We study the control of two-sided queues via different levels of information on system state. Each side is strategic and consists of patient and impatient agents (who do not tolerate any delay). We derive sufficient conditions under which the match rate is "unimodal" in the specificity of the information.

3 - Invisible Customers in Omnichannel Services

Guanling Yang, Simon Business School, University of Rochester, Rochester, NY, United States, Ricky Roet-Green

In omnichannel services, walk-in customers are physically present in the store and therefore visible to others. However, online customers are typically invisible due to the inherent nature of the online channel. This invisibility poses a significant challenge for customers' decisions about whether to wait for service. In our study, we propose two information systems to address this invisibility and the consequent information asymmetry. In the symmetric information system, customers from both channels only have access to visible queue length information. In the asymmetric information system, online customers have full queue length information, while walk-in customers only observe the visible customers in the queue. We study the strategic behavior of customers and its impact on system performance, comparing these models to the fully-observable queueing model.

4 - Dynamic Control of a Multiclass Omnichannel Production System with Applications to the Restaurant Industry

Amir Alwan, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Nasser Barjesteh

Omnichannel strategies have revolutionized production systems in various sectors, including the restaurant industry, but have also revealed challenges in demand and operations management. To address these challenges, there is growing interest in dynamic pricing for omnichannel restaurants. Motivated by these considerations, we develop a stochastic processing network model for an omnichannel production system that caters to a market of price- and delay-sensitive customers. The firm offers make-to-order (MTO) and make-to-stock (MTS) goods, available through both walk-in and online channels, at multiple predetermined quote times. The objective is to maximize long-run average profit by making dynamic pricing, production scheduling, and admission control decisions. MTO orders are subject to both earliness and tardiness costs, whereas MTS orders only face tardiness costs, alongside holding costs for the stored MTS goods. Walk-in customers have finite patience and may abandon the system if their wait time is too long. Since this problem appears analytically intractable, we consider an approximating Brownian control problem in the heavy traffic regime. We prove that the optimal policy is a two-sided barrier policy with a state-dependent drift rate. We then propose a joint dynamic pricing, scheduling, and rejection policy by interpreting this solution in the context of the original production system. Finally, we demonstrate the effectiveness of our proposed policy through a comprehensive simulation study and provide managerial insights for the dynamic control of omnichannel production systems.

ME52

Summit - 444

Spatially Distributed Service Systems

Invited Session

MSOM: Service Operations

Chair: Philipp Afeche, University of Toronto, Rotman School of Management, Toronto, ON, Canada

1 - Demand Equilibria in Spatial Service Systems

Xiaoshan Peng, Indiana University, Bloomington, IN, United States, John Carlsson, Ilya Ryzhov

A service is offered at certain locations (``facilities") in a geographical region. Customers can appear anywhere in the region, and each customer chooses a facility based on travel distance as well as expected waiting time. Customer decisions affect waiting times by increasing the load on a facility, and thus impact other customers' decisions. The service provider can also influence service quality by adjusting service rates at each facility. Using a combination of queueing models and computational geometry, we characterize demand equilibria in such spatial service systems. An equilibrium can be visualized as a partition of the region into service zones that form as a result of customer decisions. Service rates can be set in a way that achieves the best possible social welfare purely through decentralized customer behavior. We provide techniques for computing and visualizing demand equilibria, as well as calculating optimal service rates. Our analytical and numerical results indicate that, in many situations, resource allocation is a far more significant source of inefficiency than decentralized behavior.

2 - Spatial Service Network Design

Sina Hoveida, University of Waterloo, Waterloo, ON, Canada, Hossein Abouee Mehrizi, Ming Hu

This study aims to design a service network in an urban area with continuous demand to maximize social welfare. We assume that customers are sensitive to travel and wait time and seek service at a service location that maximizes their utility. First, we show that the problem of spatial service design and pricing for an urban region with strategic customers is equivalent to a bilevel design problem in which customers can be explicitly assigned to service locations. This helps us to demonstrate that the urban region can be divided into a set of connected and disjoint sub-regions such that, in each subregion, all customers seek service from the same service location. We then derive the relationship between the optimal demand rate served at each service location and the optimal capacity (service rate) of the location. We observe that the optimal service capacity follows the square root rule and show how changes in the wait and travel costs would affect the optimal demand and capacity at each location. Eventually, we characterize the relationship between the optimal service price and arrival rates and show that for locations with higher arrival rates, the optimal service price is lower if service rates are fixed. However, if the service capacity and price are determined simultaneously, the optimal service price is higher at locations with a higher optimal arrival rate.

3 - Pricing Time-Based Contracts

Tonghun Lee, University of California, Berkeley, Berkeley, CA, United States, Ozan Candogan, Yiding Feng

Digital freight brokerages connect shippers and carriers online. However, empty backhaul miles lead to inefficiency because drivers prefer cycles that start and end at the same location but struggle to find return shipments, potentially diminishing drivers' platform participation. To address this, we propose time-based contract design where the platform assigns cycles to drivers, who are compensated for their on-duty time. We use a transportation network model that conceptualizes the interactions between shippers, the platform, and drivers through a game-theoretical framework. We characterize the driver equilibrium with a potential function that reflects the unilateral deviation incentives of all drivers, and market clearing constraint ensuring that overall market conditions are satisfied. Our approach to contract design is grounded on a bi-level optimization problem, where the platform seeks to maximize its profits while adhering to the constraints imposed by the driver equilibrium. With a simple stylized network example, we observe that the platform enjoys a profit improvement after establishing contracts, even when it has access to the same set of cycles as the drivers. For general transportation networks, we explore how the level of sophistication influences the platform's contract decisions. We further address the platform's decision to expand the market on both supply and demand perspectives. Finally, we conduct empirical study with US freight data to validate our model and highlight some important managerial implications of the model.

4 - When A Platform Competes with Third-Party Sellers in Networked Markets: A Revenue Management Perspective

Hongfan Chen, CUHK Business School, Shatin, Hong Kong, Hai Wang

We consider a platform marketplace with both third-party and first-party sellers. The platform charges commissions to third-party sellers and buyers for transactions in the marketplace. Meanwhile, it also directly determines transaction prices for first-party sellers in their sales to buyers. Sellers and buyers are divided into different types with the compatibility captured by a bipartite network. Different types of sellers and buyers are heterogeneous in their cost and utility functions. Given the platform's decisions on prices and commissions, buyers/third-party sellers maximize their own payoffs from demanding/supplying products, and market-clearing conditions are satisfied in the networked market. We develop a convex optimization formulation for the platform's profit-optimal price-commission vector. Moreover, we characterize how the platform's profit-optimal price-commission profile depends on the network structure featured by complement and substitution of agents' trading relation in the marketplace. When introducing additional first-party sellers into the marketplace, we find that the widely-applied greedy strategy can perform arbitrarily bad. The platform needs to take into account the network structure in introducing the additional first-party sellers. For the fairness consideration between the platform and its market participants, we develop an efficient $(1-\epsilon)$-approximation algorithm to obtain a price-commission profile under which a fair allocation of surplus between the platform and its market participants is guaranteed in the equilibrium trades. Lastly, we apply inverse optimization to a real-world dataset to estimate the problem parameters, and provide numerical insight into the platform's revenue management strategy.

5 - Effectiveness of supply-side financial incentives in ride-hailing networks with spatial demand imbalance and strategic drivers

Uta Mohring, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Philipp Afeche, Andre Cire

When matching riders with self-interested drivers over a spatial network, ride-hailing platforms face spatial demand imbalances, which require driver repositioning to serve the total rider demand, and self-interested drivers, who strategically decide whether to participate and if so, when and where to reposition. We study the effects of supply-side financial incentives (wages) on drivers' decisions and the resulting equilibrium and evaluate their effectiveness in achieving the equilibrium under centralized repositioning.

We analyze the steady-state behavior of a ride-hailing network with general topology and general demand imbalance patterns based on a game-theoretic fluid model. We characterize the system equilibrium under decentralized repositioning for various wage policies as well as how the effectiveness of supply-side financial incentives depends on the interplay of the network's spatial configuration, wage flexibility, and the congestion-sensitivity of travel times. More precisely: (i) We identify necessary and sufficient travel time conditions for equilibrium existence. (ii) For networks with constant travel times, we show that the centrally optimal repositioning flows can always be implemented under decentralized repositioning, some wage flexibility is critical to achieve the centrally optimal capacity level and platform profit, and more limited wage flexibility yields inefficiencies in terms of drivers idling. (iii) For networks with congestion-sensitive travel times, the centrally optimal repositioning flows can generally not be implemented under decentralized repositioning, even with full wage flexibility, and decentralized repositioning implies higher capacity and average driver wage rates than centralized repositioning.

Our results provide novel managerial guidelines for the design of supply-side financial incentives under decentralized repositioning.

ME53

Summit - 445

Information Design in Supply Chains

Invited Session

MSOM: Supply Chain

Chair: Rene Caldentey, The University of Chicago, Chicago

Co-Chair: David Simchi-Levi, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Prem Talwai

1 - Information Design and Sharing in Supply Chains

Rene Caldentey, The University of Chicago, Chicago, IL, United States, Clifford Hurvich, Avi Giloni, Yichen Zhang

We study the interplay between inventory replenishment policies and information sharing in the context of a two-tier supply chain with a single supplier and a single retailer serving a Gaussian market demand. We investigate how the retailer's inventory policy impacts the supply chain's cumulative expected long-term average inventory costs in two extreme information sharing cases: (a) full information sharing and (b) no information sharing. We formulate an infinite-dimensional optimization problem whose decision variables are the MA(infinity) coefficients and interpreters them as elements of the Hardy space H2. This representation allows us to use a number of results from H2 theory to compute the optimal cost and characterize a sequence of epsilon-optimal inventory policies under some mild technical conditions. By comparing the optimal solution under full information sharing and no information sharing we derive a number of important practical takeaways. For instance, we show that there is value in information sharing if and only if the retailer's optimal policy under full information sharing is not invertible with respect to the sequence of demand shocks. Furthermore, we derive a fundamental mathematical identity that reveals the value of information sharing by exploiting the canonical Smirnov-Beurling inner-outer factorization of the retailer's orders when viewed as an element of H2. We also show that the relative value of information sharing can grow unboundedly when the cumulative supply chain costs are dominated by the supplier's inventory costs.

2 - Information Sharing in Supply Chains: A Finite Sample Perspective

Prem Talwai, MIT, Cambridge, MA, United States

We study the interplay between inventory replenishment policies and information sharing in the context of a two-tier supply chain with a single supplier and a single retailer serving a Gaussian market demand. We study the setting in which there is no information sharing between the retailer and supplier, where the supplier is further constrained to design an inventory policy using only a finite sample of past orders. We examine the interplay of sampling and approximation in the long-past asymptotics of the cumulative expected inventory cost, and design a sequence of causal, invertible ARMA policies that asymptotically achieve the optimal cost under full information sharing.

3 - Are Relay Networks Always More Expensive Than Point-to-Point Networks?

Himani Ananthula, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Milind Sohoni, Achal Bassamboo

We compare the long-run average costs of operating relay networks—including fixed costs, shipment holdover costs, and staffing costs—versus point-to-point (P2P) networks. Using data from a large trucking company, we find that the driver's likelihood of trip acceptance negatively correlates with the length of the preceding trips. Given this supply uncertainty, our model demonstrates that there exists a threshold on cost parameters above which the relay network is optimal and below which P2P becomes optimal. Additionally, under parameter uncertainty of demand, switching to a more adaptive dynamic staffing policy not only brings cost efficiency but also results in the relay network being the optimal network structure—a win-win outcome for both drivers and service providers.

4 - Planning with Supply Yield Uncertainty: On the Optimality of Linear Policies

Jan Van Mieghem, Northwestern University, Evanston, IL, United States, Riccardo Mogre

I will review work on inventory planning under demand and supply yield uncertainty and present new optimality conditions and explicit solutions for the associated single-period ``newsvendor'' model. In general, the stochastic-optimal order policy is non-linear in the starting inventory level x. The literature has proposed Linear Inflation Rules (LIR) that inflate the classic order-up-to policy. We prove that LIR are only stochastically optimal in the degenerate setting where either demand or (not and) supply yield is uncertain. Given the good performance of LIR we investigate their optimality under two robust formulations and provide bounds. We prove that LIR are robustly optimal when only the support of the demand and supply yield distributions are known. Yet we also provide the optimal distribution-free policy when the first two moments of those distributions are known, which again is non-linear. Both models provide novel, explicit order rules that may be useful for dynamic inventory control. Joint work with Riccardo Mogre (Durham U)

ME54

Summit - 446

Frontiers in Technology, Innovation, and Entrepreneurship

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Guillaume Roels, INSEAD, Boulevard de Constance, Fontainebleau, 77305, France

1 - The Role of Contextual Information in Customer Reviews: A Field Experiment on a Mental Health Platform

Guang Cheng, National University of Singapore, Singapore, Singapore, Sidika Candogan, Joel Goh, Bilal Gokpinar

Online mental health platforms can increase the accessibility of mental health care by allowing customers to select a therapist from a large pool of vetted professionals. However, customers may find it difficult to choose a therapist best suited to their needs because therapists typically list multiple specializations that tend to overlap significantly. This makes it challenging for online mental health platforms to attract new customers and retain existing ones. Motivated by this challenge, we conduct a randomized field experiment in partnership with Singapore's largest online mental health platform. In this experiment, we empirically investigate whether a simple intervention on the customer reviews--revealing contextual information about the review writers--can improve the platform's ability to attract and retain customers. We find that, relative to the control, treated customers are 42% more likely to book their first therapy session, and they book 87% more sessions in total during the experiment. We also find that customers in the treatment group spend more time searching on the platform, which at least partially explains the improvement in the platform's ability to attract and retain customers. Furthermore, treated customers who have the first therapy session book 36% more sessions with their first therapist chosen, and are twice as likely to write a positive review. Overall, our results suggest that revealing contextual information about the review writers improves the informativeness of customer reviews on mental health platforms, and it is a simple and effective way to increase successful matches and foster long-term relationships between customers and the platform.

2 - Look Before You Leap? An Operational Analysis of Hybrid Entrepreneurship

Zeya Wang, Georgia Institute of Technology, Atlanta, GA, United States, Morvarid Rahmani, Karthik Ramachandran

In this paper, we investigate the ongoing debate between fully committed entrepreneurship and the hybrid approach. Different from committed entrepreneurship, hybrid entrepreneurship involves individuals concurrently pursuing traditional employment and entrepreneurial activities. Through a model-based investigation, we analyze the trade-offs associated with these two approaches and identify the conditions that favor the hybrid approach, as well as situations where a full commitment is optimal. Furthermore, we derive the optimal allocation of effort for entrepreneurs who choose the hybrid approach. Additionally, we examine how various factors, such as the learning environment and cost of launching can impact the entrepreneur's optimal strategy.

3 - Transparent OR Not? Optimal Performance Feedback in Gamified Services

Lin Chen, INSEAD, Fontainebleau, France, Guillaume Roels

In many games or gamified services (e.g., fitness), users often receive feedback upon service completion. Such performance feedback, sometimes presented together with a goal and/or other users' scores, shapes their perception of individual performance (through prospect theory) and relative status (through social comparison). How transparent should service providers be in their disclosure of individual performance feedback to enhance user utility?

In this paper, we employ a Bayesian persuasion framework to determine the optimal information disclosure policy, taking into account whether a goal is specified and whether the other users' scores are communicated. We find that when no goal is specified but the other users' scores are communicated, full (resp., no) information is optimal when users are ahead-seeking (resp., behind-averse). When a goal is specified but the other users' scores are not communicated, the provider should reveal the exact scores to the low-performing users and only tell the high-performing users that they lie in the top range. When a goal is specified and the other scores are communicated, the optimal information policy is again only fully transparent for a subset of users: When users are ahead-seeking, it is optimal to inform users in a range containing the goal that they are in that range; whereas, when they are behind-averse, it is optimal to tell high-performing users their performance is in the top range. Our paper demonstrates how gamified service providers can enhance user utility, and thus increase the value of their service, by engineering the design of their relative performance feedback.

4 - Optimal Loyalty Program Design in a Multichannel Context

Antoine Feylessoufi, Vanderbilt University, Nashville, TN, United States, Ersin Korpeoglu, Mika Sumida

With the advent of booking platforms in the hospitality industry, consumers can now often choose between booking a room directly through the hotel or indirectly through a platform. From a hotel perspective, there is a clear trade-off. Platforms may help hotels reach a larger customer base but it comes at the expense of the profit margin due to commissions. Hotels would like to attract new customers but keep the recurring customers to book directly though their direct channel. We explore theoretically how to design an optimal loyalty program in this unique context and use a novel dataset spanning over 4 years of a group of 12 hotels to calibrate our findings.

5 - Organizational Culture, Innovation and Competitive Performance: A Multi-Level Dynamic Model

Christoph Loch, Cambridge University, Cambridge, United Kingdom, Stylianos Kavadias, Konstantinos Ladas

We develop a dynamic evolutionary model of innovation culture and its organizational effects at three levels. At the front line level, infor- mal bottom-up innovation ideas are driven by emergent interactions governed by cultural norms. As the norms and interactions cannot be monitored, management cannot control culture but only imperfectly nudge it at the second level. At the third level, the culture influences innovation performance and productivity of the firm and is thus rele- vant for the organization’s competitiveness. This approach identifies an explicit plausible pathway of how top-down strategy co-exists with implicit bottom-up driven cultural evolution of innovation.

ME55

Summit - 447

Towards Resilient and Equitable Infrastructure and Communities

Invited Session

Public Sector OR

Chair: Andrés González, University of Oklahoma, Norman, OK, 73019, United States

1 - Fairness-based resilience enhancement in network systems

Andrés González, University of Oklahoma, Norman, OK, United States

Modeling and enhancing the resilience of complex systems such as supply chain and infrastructure networks is critical for societal wellbeing. However, when studying resilience using only system-level performance metrics may lead to high inequalities, potentially affecting vulnerable regions and communities disproportionally. Thus, it is critical to develop models that not only consider diverse performance metrics at a system level, but also procure a fair distribution of resources so that risks are reduced among different sectors, and mitigation and recovery actions benefit different actors. To address this, we propose a multiobjective mathematical programming framework that integrates fairness and system-level performance, to unveil key dynamics and tradeoffs.

2 - Dynamic Mode Decomposition Analysis for Resilient Supply Chain Networks

Daniel Cabrera Giraldo, University of Oklahoma, Norman, OK, United States

This study explores the integration of Dynamic Mode Decomposition (DMD) and its Augmented variant (ADMD) with linear programming to model and optimize supply chain operations. We apply DMD and ADMD to derive transformation matrices that capture the temporal evolution of sales. These matrices are then
embedded into a linear programming model to simulate the supply chain dynamics, allowing for the endogenous forecasting of demand rather than relying on fixed
demand inputs. Our findings demonstrate the potential of combining advanced data-driven forecasting techniques with optimization models to enhance decision-making in supply chain operations, providing a more flexible and responsive approach to supply chain management.

3 - Consensus-Based Decentralized Decision-Making for Global Relocation Planning Under Climate Change

Buket Cilali, University of Oklahoma, Norman, OK, United States, Kash Barker, Andres Gonzalez, Ahti Salo

The upcoming climate crisis is a significant driver for the displacement of communities. The adverse and slow-onset climate change impacts are expected to strike people worldwide, causing displacement in great quantities; a problem that has received little attention in operations research. The characteristics of the problem – including the need for long-term planning, uncertainties about the future, and the involvement of diverse populations and multiple locations experiencing different levels of climate change impacts – require a unique approach to handle this problem. For this, we take up approaches that can assist in developing relocation plans to manage climate-change-induced movements with a long-term outlook while using the resources effectively and protecting the well-being and dignity of both displaced and receiving communities. In this study, our primary intention is to strategically plan the relocation of displaced individuals, including determining their destinations and timelines. Our secondary intention is to navigate the complex decision-making processes that arise from the involvement of various local authorities while also considering the broader global humanitarian concerns. We propose a consensus-based approach that compromises the global utilitarian goals and local interests. This approach involves using an altruistic (system-optimal) global model, self-centered (self-optimal) global models, making judgment calls, and iterative usage of a base consensus model by adjusting the decisions until proportional fairness among all participating hosts is achieved.

4 - Sheltering During the Storm: An Analysis of Decision-Making in Flash Flood Events through the Protective Action Decision Model

Paula Penagos, University of Missouri - St. Louis, MO, St. Louis, MO, United States, Trilce Encarnacion

Flash floods represent a significant natural hazard characterized by sudden onset and destructive impact, posing significant challenges to disaster management and public safety. The rapid surge of water resulting from heavy rainfall or other precipitating events can result in high mortality rates due to limited response time and the unpredictability of these occurrences. Effective risk communication and timely action are crucial in mitigating the impacts of flash floods and reducing human and financial losses. However, existing studies do not comprehensively understand individual perceptions and decision-making processes during flash floods, and communication approaches for these natural disasters remain underexplored.

This research aims to bridge these gaps by developing an enhanced communication framework tailored to manage flash flood risks in emerging economies, using the city of Santo Domingo, Dominican Republic, as a case study. This study is grounded in the Protective Action Decision Making (PADM) framework, which integrates environmental and social cues, information sources, and channels to understand individual behavioral responses during disasters. Using a Hybrid Discrete Choice (HDC) modeling within the PADM framework, this research captures behavioral responses through behavioral experiments conducted within the general population of Santo Domingo to determine individual perceptions of flash flood threats, protective actions, and stakeholder trust. The findings contribute to theoretical understanding and practical application in flash flood risk management by developing an enhanced communication framework and providing actionable guidance for stakeholders involved in disaster management efforts, ultimately contributing to the resilience and safety of communities vulnerable to flash floods.

5 - Evaluation of the impact of traditional resilience strategies in supply chains subject to disruptions in times of pandemic

Jose Navarro, University of South Florida, Tampa, FL, United States, Daniel Romero, Gina Galindo

The COVID-19 pandemic has brought unprecedented disruptions to global enterprise networks across many industries. Having resilient supply chains is essential for managing the risk of such disruptions, yet research on development of resilience based strategies during pandemics has been limited. We present a comparative analysis of resilience improvement strategies for supply chains operating during pandemics and provide recommendations for decision makers. Insights from this analysis underscore the interconnectedness of resilience strategies and emphasize the need for their diversification.

ME56

Summit - 448

MIF Paper Competition

Award Session

Minority Issues Forum

Chair: Mike Gordon, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Jessye Talley, Morgan State University, Baltimore, MD, United States

1 - Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning

Wesley Marrero, Dartmouth College, Hanover, NH, United States, Gian-Gabriel Garcia, Lauren Steimle, Jeremy Sussman

Markov Decision Process (MDP) models are commonly used tools for optimizing sequential decisions under uncertainty in medical decision making. If the parameters of an MDP satisfy certain assumptions, the optimal policy is guaranteed to be monotone. Unfortunately, these assumptions are not always satisfied. Our paper defines the price of interpretability, which measures the gap between the optimal and an interpretable policy. We assess the price of interpretability of the best-performing monotone policy and the novel class-ordered monotone policy, which preserves interpretability along user-defined state and action classes. Within the context of hypertension treatment, we demonstrate that the class-ordered monotone policy can be computed faster and achieves greater total quality-adjusted life years across a population of 66.5 million people in the US, compared to the best-performing monotone policy.

2 - Collaboration structures in integrated healthcare delivery systems: An exploratory study of accountable care organizations

Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States, Aravind Chandrasekaran, Deepa Goradia, Daniel Walker

This study explores the performance implications of collaboration structures in an integrated healthcare delivery system, namely an Accountable Care Organization (ACO). ACOs consist of providers from various stages of the care continuum (pre-acute, acute, and post-acute) that voluntarily assume collective responsibility for the quality and cost of care for a defined patient population. Studies in the healthcare operations management primarily have focused on collaboration within a single organization, shedding little light on this problem. We address this issue by exploring two distinct dimensions of collaboration: partnership scope and scale. Partnership scope measures the presence of providers from the pre-acute, acute, and/or post-acute care-continuum stages, while partnership scale measures the presence of providers within a single care-continuum stage. Leveraging a unique dataset of provider types, collaboration structures, and system-level performance for 528 Medicare Shared Savings Program (MSSP) ACOs from 2013-16, we find that synergies exist between partnership scope and scale with respect to ACO performance. We also show that the benefits of increasing partnership scope are consistent across providers and patient-level analysis. Further, we find that these benefits come at some cost, suggesting an initial cost-quality tradeoff when developing collaboration structures. Our results offer important insights into the healthcare operations management literature on designing effective healthcare delivery systems extending beyond a single organization.

3 - Content Promotion for Online Content Platforms with the Diffusion Effect

Yunduan Lin, University of California, Berkeley, Berkeley, CA, United States, Mengxin Wang, Heng Zhang, Renyu Zhang, Zuo-Jun Shen

Problem definition: Content promotion policies are crucial for online content platforms to improve content consumption and user engagement. However, traditional promotion policies generally neglect the diffusion effect within a crowd of users. In this paper, we study the candidate generation and promotion optimization (CGPO) problem for an online content platform, emphasizing the incorporation of the diffusion effect. Methodology/results: We propose a diffusion model that incorporates platform promotion decisions to characterize the adoption process of online content. Based on this diffusion model, we formulate the CGPO problem as a mixed-integer program with nonconvex and nonlinear constraints, which is proved to be NP-hard. Additionally, we investigate methods for estimating the diffusion model parameters using available online platform data and introduce novel double ordinary least squares (D-OLS) estimators. We prove the submodularity of the objective function for the CGPO problem, which enables us to find an efficient $(1-1/e)$-approximation greedy solution. Furthermore, we demonstrate that the D-OLS estimators are consistent and have smaller asymptotic variances than traditional ordinary least squares estimators. By utilizing real data from a large-scale video-sharing platform, we show that our diffusion model effectively characterizes the adoption process of online content. Compared with the policy implemented on the platform, our proposed promotion policy increases total adoptions by 49.90%. Managerial implications: Our research highlights the essential role of diffusion in online content and provides actionable insights for online content platforms to optimize their content promotion policies by leveraging our diffusion model.

ME57

Summit - Terrace Suite 1

Healthcare and Applications

Invited Session

Health Applications Society

Chair: Yangzi Jiang, The Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A, China, People's Republic of

1 - Estimating Treatment Effect from Observational Data Using a Hidden Markov Model

Tongqing Chen, University of Chicago, Chicago, IL, United States, John Birge

Treatment effect plays an essential role in the medical decision-making process and is a critical criterion for drug testing and development. However, with unobservable confounders, the treatment effect estimation can be invalid. In this work, we develop a model using observational data in which the confounder is partially observable to estimate the treatment effect. Moreover, we assume that observability can be correlated with the treatment and confounder which relax the assumptions posted by the existing works. Using maximum likelihood estimators and a hidden Markov Chain structure, the model is proved to be identifiable under some general conditions.

2 - Breaking the Stigma: the Role of AI in Drug Purchases for Sensitive Health Conditions

Tong Shen, University of Connecticut, Storrs, CT, United States, Chen Liang, Jing Peng, Mengcheng Guan, Jianbin Li

The integration of artificial intelligence (AI) has become increasingly prevalent across digital health platforms. Leveraging the introduction of an AI assistant for disease diagnosis and drug recommendations within a major drug category tied to sensitive health issues in an online pharmacy, we investigate how the adoption of AI affects users' purchase behaviors using a difference-in-differences design. We find that the adoption of the AI assistant leads to a significant increase in users' purchases in this drug category by 48.8%, which can be delineated into a 50.7% rise in the purchase of drugs recommended by AI and a 46.3% uptick in the purchase of drugs not recommended by AI. Furthermore, we find that the positive effect of adopting the AI assistant is stronger for users with less experience in related drugs or higher privacy concerns. This finding is noteworthy because it shows that the AI assistant is particularly effective for those who are usually not proactive in seeking healthcare due to limited health literacy or high privacy concerns. Finally, we explore the underlying mechanisms and reveal that the increase in user-initiated drug searches explains about 58% of the increase in the purchase of drugs recommended by AI, and nearly all of the increase in the purchase of other drugs. Our results highlight the potential of AI in improving healthcare access in sensitive medical contexts.

3 - Online Doctor-Patient Endogenous Matching: Theoretical and Empirical Evidence from An Online Consultation Platform

Xiuwei SONG, University of Washington, Seattle, WA, United States, Zhijun Yan, Yong Tan, Lin Jia

Many market outcomes are generated through endogenous matching between parties, such as students deciding which schools to apply to and schools deciding which students to admit; executives choosing which companies to join, and companies deciding which executives to hire, etc. The online healthcare(OHC) market also constitutes an endogenous matching market where patients and doctors mutually select each other based on observable characteristics. We considered the impact of both vertical and horizontal factors on the matching outcomes, where vertical factors indicate the medical capabilities of both doctors and patients, and horizontal factors represent the medical types of both parties. We collected data from a leading OHC platform and employed a two-sided matching model to explore how intrinsic and extrinsic characteristics influence matching outcomes. Our results suggest that the medical capabilities and types of doctors and patients exhibit positive assortative matching. However, while the interaction terms for patient's medical capabilities and types show positive assortative matching, those for doctors display negative assortative matching.

4 - Matrix Bandit with Risk

Kunzhu Xie, Purdue University, West Lafayette, IN, United States, Wanning Chen

Multi-armed bandit (MAB) algorithms are commonly used in clinical trials to dynamically allocate trial drugs to patients based on real-time data. These algorithms aim to identify optimal drugs and dosage (arms) and increase overall medical benefits (rewards), such as inflammation reduction. Traditional MAB algorithms focus on maximizing rewards but often overlook harmful side effects (risks), like pain. Recent literature has explored incorporating safety constraints into the MAB framework to control the frequency of administering drugs and dosages with intolerable risk levels. However, existing methods would suffer from long burn-in periods and high exploration regret when the number of arms is large. This issue frequently arises in two-sided clinical trials, where we need to determine the optimal dosage combination for two drugs. Therefore, we propose a two-phase bandit algorithm, Matrix Bandit with Risk (MBR), leveraging the low-rank structure of both reward and risk matrices. MBR first identifies a subset of potentially optimal arms using low-rank estimations, then applies a doubly optimistic strategy based on UCB to this subset to select one arm at each time step. Simulations show that MBR has lower regret and effectively controls the number of safety violations compared to existing methods.

ME58

Summit - Terrace Suite 2

Enhancing Resilience: Strengthening Healthcare Systems for Preparedness and Response

Invited Session

Health Applications Society

Chair: Su Li, Texas A&M University, College Station, 77840

1 - Makeshift Hospital Utilizations During COVID 19 Pandemic: An Efficiency Problem

Eyyub Kibis, Montclair State University, Montclair, NJ, United States, Musa Caglar, Ali Dag, Anton Ivanov, Ozlem Cosgun, Serhat Simsek

In this study, we develop a compartmental dynamic MIP model to investigate the transmission of the COVID-19 pandemic within a community. We propose policies that will help state governments control the pandemic and reduce patient density at overcrowded hospitals. Our model predictions closely fit the real outbreak data and suggest that the deployment of makeshift hospitals as post-treatment facilities will result in the most efficient use of resources to control the pandemic.

2 - Resilience and Adaptation: Assessing the Effects of COVID-19 on Hospital Efficiency and Quality

Utkarsh Verma, Schulich School of Business, York University , Toronto, ON, Canada, Adam Diamant, Raha Imanirad, Amol Verma, Fahad Razak

This study examines the impact of COVID-19 on various aspects of hospital operations, including clinical efficiency, technical proficiency, and experiential quality. Utilizing a dataset of over 250,000 patient admissions from 15 hospitals in Ontario, Canada, we employ data envelopment analysis (DEA) to assess hospital performance based on variables such as readmission rate, mortality rate, length-of-hospital stay, the number of ED visits and surgeries, and wait time metrics for various inpatient services. We then use bootstrapped DEA scores as dependent variables in a fixed-effect, difference-in-differences regression model with heterogeneous treatment times, to explore the causal relationship between the onset of COVID-19 and a hospital's operational scores. Our findings reveal a statistically significant negative impact on clinical efficiency which can be attributed to the COVID-19 pandemic, whereas technical proficiency and patient experience exhibit resilience amidst well-documented operational challenges. The consistency of these results across various robustness checks and alternative specifications underscores their reliability. Consequently, this study deepens our understanding of the implications that COVID-19 had on hospital operations and provides valuable insights for healthcare management strategies in the face of such disruptions.

3 - Exploring the Relationship Between Health Information Exchange, Experiential Quality, and Medical Malpractice: a Hospital Perspective

Sukrit Pal, Iowa State University, Ames, IA, United States, David Cantor, Anand Nair

This study investigates the relationship between hospital participation in Health Information Exchange (HIE) and medical malpractice lawsuits, and explores the moderating effect of experiential quality on this association. Through a comprehensive analysis of healthcare facilities and court (federal + state) systems, we found that hospitals engaging in HIE tend to experience higher rates of medical malpractice lawsuits. However, our research reveals that hospitals with a strong focus on experiential quality demonstrate a mitigating effect, effectively reducing the adverse impact of HIE on malpractice occurrences. By highlighting the significance of experiential quality in healthcare settings, this study contributes to understanding the intricate dynamics between information exchange initiatives and patient outcomes, offering insights for policymakers and healthcare administrators to enhance quality of care and mitigate legal risks.

4 - Evaluating U.S. Drug Resistance Surveillance Policies: a Dynamic, Multi-Simulation-Optimization Approach

Xuecheng Yin, Oklahoma State University, Tulsa, OK, United States, Dursun Delen, Yue Yuan, Esra Büyüktahtakın Toy

The United States' drug resistance surveillance system provides guidance on which antibiotics are used for the first line treatment of infectious diseases, such as gonorrhea. The system changes the empiric therapy of gonorrhea when the prevalence of resistance to the antibiotic exceeds a certain threshold (e.g., 5%). However, it is unclear how providing first-line treatment guidance based on each region independently instead of changing the first-line therapy for all regions together would impact the system. To address this question, we developed a Dynamic Optimization Multi-Simulation Stochastic Model to project the burden of gonorrhea and optimize the surveillance allocation over multiple regions. We compared several proposed policies for their influence on gonorrhea transmission, cost, and human life quality. Our findings indicate that decreasing the frequency of making decisions will not significantly influence the number of infections. However, decreasing the frequency of making decisions leads to a higher number of infections with resistance to the first-line therapy. Furthermore, switching to a new drug for each region independently leads to a lower number of infections, but increases the cost and decreases individual's life quality in the healthcare system under the current 5% switch threshold. The results will improve if a higher threshold is selected (e.g., 9%), which highlights the importance of optimizing the switch threshold and the increasing frequency of making decisions. Our model is flexible and can be extended to study other diseases with drug resistance surveillance to provide guidance on both interventions and treatment.

5 - Social Network Activation: An Application in Reducing Smartphone Overuse

Shaokang Yang, Virginia Tech, Blacksburg, VA, United States, Idris Adjerid, Jiayi Liu

This paper explores smartphone addiction through the lens of social network analysis. Using data from multiple longitudinal studies that track individual relationships and smartphone interactions, we provide empirical evidence that activating specific parts of one's social network affects smartphone usage patterns. Our results highlights the potential of strategic network interventions in managing digital addiction.

ME59

Summit - Ballroom 1

Markov Lecture

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Various Incarnations of Matrix Completion

Devavrat Shah, Massachusetts Institute of Technology, Cambridge, MA, United States

The objective of matrix completion is to estimate or complete an unknown matrix from its partial, noisy observations. Since its introduction as a model for recommendation systems in the 1990s, it has been central to advances in machine learning, statistics, and applied probability. In this talk, I will discuss a few incarnations of it that arise in the context of time-series analysis, causal inference, reinforcement learning and empirical risk minimization.

2 - Discussant

Christina Yu, Cornell University, Ithaca, NY, United States

Discussant

ME60

Summit - Ballroom 2

Amazon Geospatial Planning

Invited Session

The Practice Section of INFORMS

Chair: Liron Yedidsion, Amazon, Redmond, WA, United States

Co-Chair: Dipal Gupta

1 - Optimizing Van Loading and Dispatching Operations in Amazon's Last Mile Delivery Stations

Yaniv Mordecai, Amazon, Bellevue, WA, United States, Liron Yedidsion, Rohit Malshe

Amazon Logistics (AMZL) ships packages from hundreds of delivery stations (DS). Every day, each DS loads and dispatches scores of vans to deliver packages to customers. AMZL has been dispatching in batches: vans are grouped by sectors of the station's jurisdiction footprint, and all the vans to each sector are launched in a single wave. This method has several operational advantages, but it also results in driver frustration due to the loss of precious minutes, which could make a difference on the road for safer, less stressed shifts. The transition to a different van launching method is not trivial and requires significant changes to AMZL's modus operandi, and must be backed by sound science. We have tackled this problem from both the business and science perspectives, studied several conceptual and physical design alternatives, and built a robust simulation model to assess the impact each layout can have on both the van launch phase and on upstream logistics inside the DS. We have shown that our suggested design and operational method can conservatively save about 40% of the wave dispatch time, which results in a few additional on-road minutes per route. In this talk, we present the various business, engineering, operations, and science considerations and how they interact with each other. If implemented correctly, the proposed van loading method can revolutionize Amazon's Last Mile operations and yield significant benefits in driver safety, sentiment, and ultimately also performance.

2 - Geospatial optimisation at Amazon

Dipal Gupta

3 - Last mile drivers capacity scaling

Mahdieh Allahviranloo, Amazon, City university of New York, New York, NY, United States, Jin Ye

Capacity planning for last mile drivers is critical for operational efficiency in Amazon last mile planning. Over-solving for capacity can lead to extra headroom in resources. On the other hand, under-solving for capacity results in increased costs due to unmet customer promises and the potential outsourcing of packages. Each additional capacity lever comes with associated challenges and limitations on when and how the additional capacity can be deployed optimally. Carefully chosen capacity levers have the potential to significantly reduce costs and enhance customer service. In this paper, we present a multi-objective optimization model where its objective function takes into account the costs associated with capacity levers, as well as overage and underage costs for delivery service providers. The constraints are set to represent ongoing challenges in adjusting needed capacity for Amazon deliveries. The results of the model, designates the optimum capacity scaling strategy along with the optimum deployment timeline.

4 - Time Series Ensemble Forecasting Models

Abhilasha Katariya, Amazon, Bellevue, WA, United States, Eliot Hijano

We propose a comprehensive framework for improving time series forecasting in logistics by integrating clustering techniques, transfer learning, and ensemble modeling. We introduce a novel clustering method to group similar time series data into clusters, forming a cluster time series model, and employ transfer learning within a hierarchical structure of time series models. Additionally, we extend the ensemble modeling by integrating complementary models like physics-based or off-line models. We demonstrate the effectiveness of our approach in addressing operationally constrained metrics in logistics, offering enhanced forecasting accuracy and enabling informed decision-making.

ME61

Summit - Ballroom 3

INFORMS 2035: Where are we Going?

Panel Session

Committee’s Choice

Co-Chair: Anne Robinson, Robinson Insights, Ottawa, ON, Canada

1 - Moderator Panelist

Anne Robinson, Ottawa, ON, Canada

2 - Panelist

Ramayya Krishnan, Carnegie Mellon University, Pittsburgh, PA, United States

3 - Panelist

Margery Connor, Chevron, San Ramon, CA, United States

4 - Panelist

Lauren Czerniak, Grainger, Pittsburgh, MI, United States

5 - Panelist

Elena Gerstmann, INFORMS, Catonsville, MD, United States

6 - Panelist

Turgay Ayer, ISyE Georgia Tech, Atlanta, GA, United States

na

ME62

Summit - Signature Room

Measuring the Efficacy of Amazon’s Recommendation Systems

Invited Session

TutORial

Chair: Julie Ivy, University of Michigan, Ann Arbor, MI, United States

1 - Measuring the Efficacy of Amazon's Recommendation Systems

Ozalp Ozer, Amazon, Seattle, WA, United States, A. Serdar Simsek, Xiaoxi Zhao, Ethan Dee, Vivian Yu

Amazon's Fulfillment By Amazon (FBA) program provides assistance to Selling Partners ("sellers," for short) in the form of information sharing, recommendations guiding seller actions (e.g., restock quantity recommendations, excess inventory recommendations), and delegated actions (e.g., automated removals of aged inventory). Amazon's vision is to help sellers make better decisions and achieve better business outcomes.
In this tutorial, we consider the sophisticated optimization models Amazon employs to generate recommendations. For example, if a seller has excess inventory, Amazon recommends actions to increase their sell-through rate, such as creating a sale or Sponsored Product ad. We demonstrate how we measure the efficacy of these recommendation systems on seller-product outcomes (e.g., revenue, units shipped, and customer clicks on product listings, or "glance views"). Measuring such outcomes is a causal inference problem because we only observe each seller-product's "factual" and not their "counterfactual" outcome. We employ causal machine learning methodologies such as double machine learning, causal forest, and doubly-robust forest to separate selection bias from a comparison of "treatment" and "control" sellers. For example, we find that aligning with the restock and excess inventory recommendations, on average, improves several seller-salient outcomes. We also present methods for measuring heterogeneity in the efficacy of these recommendations across seller and product segments, and estimate personalized benefits for each seller-product. Finally, through A/B testing, we find that sharing quantified efficacy information with sellers increases their adoption of Amazon recommendations. Sellers are responding to this messaging, and the duty to them is to rigorously identify causal estimates.

ME63

Regency - 601

Algorithmic and Economic Analysis on Digital Platforms

Invited Session

Information Systems

Co-Chair: Yifan Yu, University of Texas at Austin, Austin, TX, United States

1 - Optimal Update Releasing Strategies Using Multi-Armed Bandits and Review Contents

Mingrui Zhang, University of Denver - Daniels College of Business, Seattle, WA, United States, Yingda Lu

This paper presents a novel approach to optimize game update release strategies by leveraging multi-armed bandits and sentiment analysis of game reviews. We analyzes the content of recent game reviews to determine the optimal type of update to release that maximize the averga positive reveiw rate after the release.

2 - Optimal Ranking Algorithm on Online Health Forum

Zoey Hao, University of Pennsylvania, Philadelphia, PA, United States

As online health forums have been increasingly important to help patients with their concerns on their heath condition in a timely manner, the importance of having effective ranking algorithms that can significantly impact the quality of healthcare information dissemination, user engagement, and even patient outcomes is crucial. Previous work studying ranking algorithms usually consider a series of factors, including relevance, popularity, timeliness, and source credibility. In the context of healthcare forums, the traditional factors remain important but are often insufficient. In this paper, we present an empirical framework for creating a forward-looking ranking algorithm that can boost long-term benefits of platform users.

3 - From Content to Action: Identifying Participatory Engagement in National Statistical Systems

Winry Liu, University of Texas at Austin, Austin, TX, United States

The pursuit of Sustainable Development Goals (SDGs) necessitates robust and inclusive statistical systems. This study develops and validates indicators of participatory engagement in national statistical activities, exploring both theoretical and empirical contributions. Leveraging advanced natural language processing (NLP) techniques, we efficiently identify participatory elements from textual data on official websites. Our research investigates the correlation between these participatory indicators and statistical engagement levels, examining how enhanced participation impacts sustainable development. By integrating additional data, we provide a comprehensive model for assessing participatory practices. The findings offer valuable insights for policymakers and national statistical offices, fostering a deeper understanding of the role of participatory engagement in achieving sustainable development.

4 - Navigating the Realm: The Impact of Virtual reality on Sellers' Revenue and Bidders' Tactics in Online House Auctions

Zhenbin Yan, School of Economics and Management, Tongji University, Shanghai, China, People's Republic of, Zhongyun(Phil) Zhou

Virtual reality (VR) is a promising visual technology that may transform product presentation into a 3D and interactive approach. Even though VR's influences have been widely discussed in e-commerce, most prior literature examines VR's impact on consumer behavior in selling inexpensive and small-size goods such as books and clothing using posting price. Online house auctions in our context significantly differ from prior works in terms of product type, sales mode, and research perspective. Specifically, online consumers show higher quality concerns over valuable products and are more difficult to evaluate products with rich spatial details. Besides, competitive bidding in online auctions allows potential buyers' communication through submitting bids strategically, which is distinct from consumers' independent decision-making in posting price. Given the uniqueness of online house auctions, we follow an economic perspective and conduct empirical analyses with a unique dataset to learn how VR affects sellers' revenue and bidders' tactics. Estimation results show that VR can increase houses' final transaction price and benefit sellers to obtain higher revenue. The positive effect of VR is greater for houses with higher product quality. Additional analyses reveal that VR may make buyers bidding strategically. On one side, VR facilitates buyers to place additional jump bids. On the other side, VR reduces the ratio of late bids in the whole bidding process. Both of them contribute sellers to obtain higher revenue. Our findings add to the literature on VR, the online real estate market, online auctions, and bidding strategies and provide implications for stakeholders in e-commerce.

ME64

Regency - 602

Machine Learning and Social Media Analytics

Invited Session

Social Media Analytics

Chair: Zhiyu Zeng, Washington University in St. Louis, St. Louis, MO, United States

1 - Which Model to Choose for E-Commerce Platforms in Emerging Markets? Centralized Model or Decentralized Model

jingyi peng, University of Science and Technology of China, Hefei, China, People's Republic of, Feng Yang, Zihao Zhang

E-commerce is booming and is increasingly becoming the mainstream shopping method worldwide, giving rise to a variety of e-commerce platforms. However, the performance of operational models varies across different environments. Undoubtedly, Amazon stands as the leader among e-commerce platforms, which we refer to as a centralized platform that requires both buyers and sellers to conduct transactions within the platform. A similar platform is Taobao, the leading e-commerce platform in China. In addition, another operational model has achieved success in recent years—Software as a Service (SaaS), represented by independent website-building platforms such as Shopify and WooCommerce. There are over a dozen of these types of platforms globally and SaaS platforms are frequently featured in many lists of e-commerce platforms. SaaS platforms enable merchants to create their own websites for selling goods, known as the decentralized model.

Throughout the development of e-commerce, countless platform operational models have emerged and faded. Some barely survive without scaling up, while others gradually dominate. We know that adapting to the times is key to success.

Therefore, we hope to delve into discussions on which market environments are most conducive to the development of specific platform operational models: For emerging markets, analyze which operational models best suit their conditions; For competitive markets, examine how platforms with different operational models should strategize to gain more profit and market share.

Furthermore, we can conduct a deeper analysis of the interactions between platforms, merchants, and buyers within different operational models, providing interesting findings and recommendations.

2 - Tweets to Citations: Understanding the Role of Social Media Influencers in AI Research Visibility

Iain Weissburg, UCSB, Santa Barbara, CA, United States, Mehir Arora, Xinyi Wang, Liangming Pan, William Wang

As the number of accepted papers at AI and ML conferences reaches into the thousands, it has become unclear how researchers access and read research publications. In this paper, we investigate the role of social media influencers in enhancing the visibility of machine learning research, particularly the citation counts of papers they share. We have compiled a comprehensive dataset of over 8,000 papers, spanning tweets from December 2018 to October 2023, alongside controls precisely matched by 9 key covariates. Our statistical and causal inference analysis reveals a significant increase in citations for papers endorsed by these influencers, with median citation counts 2-3 times higher than those of the control group. Additionally, the study delves into the geographic, gender, and institutional diversity of highlighted authors. Given these findings, we advocate for a responsible approach to curation, encouraging influencers to uphold the journalistic standard that includes showcasing diverse research topics, authors, and institutions.

3 - Dynamic Content Adjustment For Education Platform: Learning and Maximizing Engagement

Lingfei Zhong, The University of Hong Kong, Hong Kong, Hong Kong

A crucial problem for education platforms is how to strategically design content displayed in the app to maximize the engagement of users. In particular, for those education platforms that offer courses to users, such as Coursera and Duolingo, the design of the content taught to users could significantly affect users' activity and engagement. We focus on the design of sequencing of the levels of difficulties of the different lessons in a course. The difficulty might induce users' passion to conquer new materials, but it might also deter users' confidence, leading to users' quitting. The design of sequencing of the levels of difficulties and its effects could become very complicated even in a deterministic scenario where users' behaviors are well specified. We find that challenging content often deters users instead of encouraging users, and that dynamically adjusting course content too actively might hurt users' engagement. Studying the deterministic case might not be able to provide enough insights for course planners because users' transitions from one learning state to another learning state are highly dependent on users' innate features and users' progression as they study a course. We then propose an algorithm to consider users' innate features, progression in the course, and potential deviation from their learning behaviors, and incorporate the contextual information into the course decision-making process. We hope this algorithm might provide more practical guidance to education platform designers on how to strategically design sequencing of the levels of difficulties for content.

4 - Leveraging Machine Learning to Analyze Social Media Data: Findings from a Pilot Study

Nastaran Khalili, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, United States, Fatima BarraganHerrera, Blair Coleman, Jill Settle, Sara Eggers, Reza Kazemi-Tabriz, Roberto Delhy

This paper describes an innovative application of machine learning techniques to automate the analysis of cannabis-related discourse extracted from a commercially available social listening platform. Using a de-duplicated, manually analyzed and tagged dataset, we employed the BERT (Bidirectional Encoder Representation from Transformers) model to develop a classification algorithm that replicates the nuanced review and categorization process performed by expert qualitative analysts and generate a relevancy prediction for posts that have not been coded. Through hyperparameter tuning, our approach achieved a recall of 0.88 in predicting relevance. Subsequently, we utilized natural language processing (NLP) techniques to extract themes and topics, as well as medical and health-related conditions around people's experiences mentioned within the social media data and clustered them accordingly. Our study highlights the efficacy of this approach in efficiently analyzing social media data, which can minimize the burden of manual coding of these large, unstructured data and offer insights into publicly posted experiences, behaviors, and perceptions as a complement to traditional qualitative analysis methods.

ME65

Regency - 603

Managing Industry Advisory Boards

Panel Session

Education Outreach

Co-Chair: Ken Murphy, Merage School - UC Irvine, 4293 Pereira Drive, Irvine, CA, 92697, United States

1 - Moderator Panelist

Ken Murphy, Merage School of Business, UC Irvine, Irvine, CA, United States

2 - Panelist

James Cochran, The University of Alabama, Tuscaloosa, AL, United States

The growth of academic programs in analytics has increased the responsibility and complexity of faculty engagement with industry counterparts. In this panel academic faculty and professionals will share their experience in establishing, managing, and participating in successful industry advisory boards. Issues discussed will include setting board goals, generating industry projects, creating student internships, finding clinical faculty, and managing stakeholder expectations. The question of charging a fee for board membership will also be considered.

3 - Panelist

Jeffrey Camm, Wake Forest University, Winston-Salem, NC, United States

4 - Panelist

Matthew Bailey, Bucknell University, Lewisburg, PA, United States

5 - Panelist

Robin Lougee, KnitWell Group, Yorktown Heights, NY, United States

ME66

Regency - 604

Human-AI Connection in Digital Platforms

Invited Session

Artificial Intelligence

Chair: Chen LI, City University of Hong Kong, Hong Kong, N/A

Co-Chair: Zhiya Zuo, City University of Hong Kong, Kowloon Tong, N/A

1 - Who Moved My Cheese? Algorithm Reactance in Online Investment Communities

Chen LI, City University of Hong Kong, Hong Kong, China, People's Republic of, Zhiya Zuo, Weiquan Wang, Qiang Ye

Algorithms are being increasingly adopted in online investment communities (OICs) to generate algorithmic predictions of stock performance for OIC content consumers (i.e., investors). Nevertheless, such predictions may pose an identity threat to the OICs' key content producers (mainly consisting of non-professional human analysts), who analyze and forecast stock performance through their content generation in OICs (i.e., analysis articles). We investigate how analyst content generation is adapted in response to the introduction of algorithmic predictions in OICs. Drawing on identity control theory and coping theory, we theorize analysts' coping responses under algorithmic identity threat as an algorithm reactance effect, which manifests in two forms, namely, algorithm differentiation and self-elevation. We further elaborate on the contingent role of algorithm performance and analyst experience in influencing the intensity of the reactance effect. Applying a regression discontinuity in time design, we empirically tested the theorized reactance effect using a unique dataset from Seeking Alpha, an OIC that rolled out algorithmic predictions in May 2019. We found that analysts exhibited convergent differentiation towards accurate algorithmic predictions, while exhibiting divergent differentiation when algorithms performed poorly. Moreover, analysts demonstrated growing productivity and portfolio diversity (i.e., self-elevation), which was found to be negatively moderated by analyst experience. Our study reveals a novel, nuanced algorithm reactance effect in a human–algorithm interaction context and offers managerial implications for next-generation algorithm-empowered OICs.

2 - Investor Engagement In Algorithm-Empowered Online Investment Communities

Zhiya Zuo, City University of Hong Kong, Kowloon Tong, Hong Kong, Chen LI, Weiquan Wang, Qiang Ye

Algorithms are increasingly integrated into online investment communities (OICs) to generate quantitative stock predictions for investors. Albeit a seemingly novel feature, it remains unknown whether and how such predictions reshape investor engagement with peer-generated stock analysis. Drawing upon information foraging theory, we theorize the effects of algorithmic predictions on investors' engagement depth (i.e., the intensity of investors' interactions with analysis articles) and engagement breadth (i.e., the diversity of investors' interactions across various stocks). We collected data from Seeking Alpha, an OIC that introduced algorithmic predictions in May 2019. Using a panel dataset of 43,680 single-ticker analysis articles on 1,820 stocks, we implemented a regression discontinuity in time design to test our hypotheses. Our results reveal a nuanced effect on engagement depth contingent on investor experience and analyst–algorithm prediction incongruity. Meanwhile, algorithmic predictions increase engagement breadth through portfolio diversification. This study contributes to human–algorithm connection in a novel context of online communities by focusing on both depth and breadth of engagement. We further discuss managerial implications for OIC platforms and investors.

3 - The Value and Credibility Judgments of AI-Generated Summaries from Consumer Reviews

Zheyi Xu, University of South Florida, Tampa, FL, United States

Platforms have started to adopt AI-generated summaries from consumer reviews to simplify review consumption and facilitate decision-making. However, we know little about the expected value and credibility of such innovations. To examine this question, we distinguish two formats of AI-generated review summaries: narrative-based and attribute-based. We also evaluate their influence on the expected value and credibility, which are assumed to change in the same direction in prior literature. Drawing on fluency theory and feelings-as-information theory, we propose that narrative-based (vs. attribute-based) summaries generated by AI can improve both perceived helpfulness and perceived credibility. We conducted two complementary laboratory experiments to examine this question and demonstrated several key findings. First, AI-generated narratives were perceived to be more helpful than AI-generated attributes because narratives were perceived to be more fluent. In addition, AI-generated summaries were perceived to be more helpful compared to not providing them, but the provision of AI-generated attributes did not enhance perceived helpfulness. Second, perceived helpfulness and credibility did not improve concurrently because the two formats of AI-generated summaries were similar in credibility. Moreover, providing either format was perceived to be similar in credibility compared to not providing them. Third, we found that when the summaries were provided by humans, the inconsistency between the two outcomes became even greater. Our research provides important theoretical and managerial implications.

4 - Spillover Effects of Generative AI on Human-Generated Content Creation: Evidence from a Crowd-Sourcing Design Platform

Weihong Zhao, University of Maryland College Park, Collage Park, MD, United States, Wen Wang, Siva Viswanathan

This study investigates the spillover effects of generative AI on organic human-generated designs. Using extensive data from an online crowd-sourcing design platform, we examine how the adoption of generative AI influences novelty of designers' organic human-generated creations. Additionally, we examine the moderating roles of designers' AI-expertise and domain-expertise. We employ advanced deep learning models and well-established econometric methods to causally identify the impact. Our findings reveal that initial reliance on generative AI may suppress creativity, but long-term use enhances it. A key set of findings relate to the role of designers' AI-expertise and domain-expertise. We find that AI-expertise as well as domain-expertise, independently and jointly, contribute to the novelty of organic designs. Designers with dual expertise, both AI and domain, benefit the most, highlighting the complementarity of these skills. Moreover, our prompt analysis shows that designers with either AI-expertise or domain-expertise use lengthier, more detailed prompts, leading to improved outcomes. Prompts used by designers with domain-expertise are more distinctive than those used by designers with AI-expertise. Overall, our study carries implications for crowd-sourcing platforms and policymakers, highlighting the economic value of generative AI in the design industry and its role in fostering creativity.

5 - Keep The Positive in The Loop: A Field Experimental Study of The Impact of Integrating Solicited Consumer Feedback in Content Recommendation Systems

Yilin Li, Peking University, Beijing, China, People's Republic of, Changrong Xiao, Chong Wang, Sean Xu, Jiayin Zhang

Recommender systems mostly rely on observed user engagement behavior for training and optimization. Yet, behavioral feedback has known biases in capturing consumers' preferences, which necessitate the inclusion of actively solicited user feedback in recommendation systems. We report findings from a field experimental study of the impact of integrating solicited consumer feedback in content recommendation systems. In collaboration with a significant social media platform, a pop-up survey module was introduced to actively collect content consumers' feedback (satisfied/dissatisfied/uncertain). The solicited consumer feedback was used to adjust video recommendations following either the "boosting" or "filtering" strategy. The results show that the two methods for incorporating solicited consumer feedback have distinctive impacts. The boosting strategy increased content diversity and user activity, especially for new users, while the filter strategy decreased content diversity and user engagement. This research reveals the nuances in the design of the integration of "human" and "algorithm" in "keep-human-in-the-loop" recommender systems.

6 - Keep the Positive in the Loop - a Field Experimental Study of the Impact of Integrating Solicited Consumer Feedback in Content Recommendation Systems

Chong Wang, City University of Hong Kong, Hong Kong, Hong Kong

AI recommenders mostly rely on passively observed user engagement behavior (e.g., clicks, likes, comments) for training and optimization. Yet, passive behavioral feedback has known biases in capturing consumers' preferences, which necessitate the inclusion of actively solicited user feedback in recommendation systems. We report findings from a field experimental study of the impact of integrating solicited consumer feedback in content recommendation systems. In collaboration with a significant social media platform, a pop-up survey module was introduced to actively collect content consumers' feedback (like/dislike the content). The solicited consumer feedback was used to adjust personalized recommendations following either the "boosting" or "filtering" method. The results show that the two methods for incorporating solicited consumer feedback have distinctive impacts. The boosting method increased content diversity and user activity, especially for new users, while the filter strategy decreased content diversity and user engagement. This research reveals the nuances in design the integration of "human" and "algorithm" recommendations in "keep-human-in-the-loop" recommender systems.

ME67

Regency - 605

Advancing AI: From Classification to Molecular Predictions

Contributed Session

Contributed

Chair: Ayse Dogan, UIUC, 2111 Hazelwood Dr, URBANA, IL, 61801, United States

1 - Comparison of the Physiological Signals to Explain the Personal Anxiety Level

Ayse Dogan, University of Illinois at Urbana-Champaign, URBANA, IL, United States, Richard Sowers, Manuel Hernandez

Anxiety has emerged as a significant health concern in recent times, prompting extensive research into its rapid detection and assessment. Physiological data, often collected through wearable sensors, play a crucial role in this endeavor. However, challenges remain in gathering this data and accurately interpreting the stress levels individuals experience. This study aims to address these challenges by comparing various physiological datasets and their correlation with self-reported stress levels using the STAI6 survey. We analyzed Electrodermal Activity (EDA), Blood Volume Pulse (BVP), Heart Rate Variability (HRV), Temperature (TEMP), and Speech datasets from participants during a TSST, evaluating their effectiveness in explaining self-reported stress levels. Our findings provide insights into the relative utility of these physiological measures in stress assessment.

2 - Training Graph Neural Networks with Consistency Regularization for Molecular Property Prediction

Jongmin Han, Sungkyunkwan University, Suwon, Korea, Republic of, Seokho Kang

Although graph neural networks (GNNs) have proven to be powerful in various molecular property prediction tasks, GNNs often suffer from poor predictive performance when trained with small training datasets. Data augmentation is a viable option to alleviate data insufficiency. However, perturbing molecular graphs may potentially alter their properties, thereby negatively affecting predictive performance. In this study, we propose a consistency regularization method to better utilize data augmentation during the training of a GNN. Various molecular graph augmentations, including atom masking, atom deletion, bond deletion, and substructure removal, are used to obtain multiple augmented views of a molecular graph. By introducing a consistency regularization loss into the supervised learning objective, we encourage the GNN to ensure that the predictions for the augmented views of a molecular graph are close to the prediction for the same graph. Through experimental evaluation, we demonstrate that the proposed method outperforms existing methods that leverage data augmentation, especially when the training dataset is smaller.

3 - Lumped Deep Learning Algorithm Based on Pseudo Inverse Techniques&Mdash;a Prediction Model for Urban Development

Tachia Chin, Zhejiang University of Technology, Hangzhou, China, People's Republic of, Zhibin Zhang, Weihua Chieng

This paper proposes a new machine learning algorithm, called the lumped deep learning algorithm, using the pseudo-inverse and its residue to backpropagate and train the neuron weights. The algorithm relies on the matrix operators, which can be easily implemented with the existing libraries such as EIGEN or LINPACK with high computational efficiency. Needed no gradient information, the loss function and the error function can be defined in many ways without causing much burden on the computational complexity. Since the pseudo inverse is intrinsically a least square method, the loss function may be chosen as the maximum standard deviation error among the training examples. Comparing the average error with the maximum error, we can justify whether the size of a neural network is proper for a specific machining learning task. The pseudo-inverse and its residue is backpropagated with low learning rates to make the annealing process of neural weights causing all examples to gradually agree. When the number of layers and neurons can be determined by the number of independent training examples, the neural network can be easily configured without much trial effort. The city development prediction is used as an example, which involves the prediction of the population growth and the regional GDP growth. Despite the data discrepancy for the pseudo-inverse method, the lumped deep learning can still integrate different categories of data and consolidate them to perform the prediction. The result is compared with the conventional regression analysis, a significant improvement was found and discussed in the paper.

4 - A Comparative Analysis of Number of Alternatives Generated by OpenAI and Cohere Language Models for Technology-Related Decision Making

Minoo Ahmadi, University of Southern California, Los Angeles, CA, United States

This research conducts a comparative analysis of the OpenAI and Cohere language models, specifically evaluating their capacity to generate decision-making alternatives for technology-related queries. Utilizing authentic queries from the Quora platform, two distinct batches of 50 questions each—focused on technology options and comparisons—were analyzed. Questions were manually curated to ensure clear decision-making contexts with a strong technological focus, categorized into "Career Choices" and "Technology Comparisons." Both models were tasked with generating alternatives for each question, guided by a standardized definition of "alternative." The analysis revealed that Cohere consistently generated more alternatives than OpenAI across both batches, suggesting that Cohere's capabilities in practical application and text analysis may enhance alternative generation in decision-making scenarios. This indicates a potential for Cohere to better support human decision-making by providing a broader range of options. These findings highlight the significant role of language models in enhancing decision support systems and suggest that leveraging such technologies can improve the quality of decisions in technology-driven contexts. Future research will further explore the quality and diversity of the alternatives generated and how these models can be integrated into comprehensive decision-support frameworks.

5 - Managing Noise and Lack of Stationarity with AI

Irene Aldridge, Cornell University, New York, NY, United States

Six key problems are known to plague traditional econometric analysis and much of pure machine learning: 1) overfitting, 2) inability to deal with noise in the data, 3) difficulties dealing with missing observations, 4) data non-stationarity, 5) researcher ability and creativity, and 6) researcher personal biases. In this paper we show that Artificial Intelligence (AI), when applied in a knowledgeable and correct way, helps researchers deal with the six problems in a fast and efficient manner, resulting in robust population inferences.

ME68

Regency - 606

Innovations in Machine Learning for Complex Data Analysis

Contributed Session

Chair: Hyungu Kang, Sungkyunkwan University, Jangan-gu, Suwon-si, Korea, Republic of

1 - Industrial anomaly detection based on foreground-focused diffusion

Hyungu Kang, Sungkyunkwan University, Suwon, Korea, Republic of, Seokho Kang

Recently, diffusion models have emerged as effective tools for unsupervised anomaly detection by leveraging their reconstruction capabilities. To detect anomalies in a query image, noise is incrementally added over multiple time steps, followed by denoising using the diffusion model across the same number of time steps. The reconstruction error serves as the anomaly score. Considerable research efforts have been dedicated to developing effective diffusion models for improved anomaly detection performance. These methods apply the diffusion process to the entire regions of an image. In typical industrial anomaly detection tasks, anomalies can only appear in the foreground region of an image where the target object lies. Thus, applying the diffusion process to the background region is unnecessary. Additionally, the location of the target object can provide important information for the reconstruction. In light of this, we propose a foreground-focused diffusion method that applies the diffusion process exclusively to the foreground region. We separate the foreground and background regions in images using the segment-anything model. During both the training and inference phases, we apply the diffusion process only to the foreground region by masking the background region at each time step. The proposed method can be used as an add-on to any existing diffusion-based anomaly detection method to improve performance. We demonstrate the effectiveness of the proposed method through experiments conducted on manufacturing and medical image datasets.

2 - Comparing various feature selection methods to identify crucial disease-associated biomarkers

Shiva Afshar, Emory University, Atlanta, GA, United States, Eric Dammer, Erik Johnson

Feature selection is critical in machine learning because it enables models to focus on the most important information. While various feature selection techniques exist, many rely on basic methods like filtering, wrapping, embedding, and dimensionality reduction. These methods may be impacted by noise and outliers and lack the ability to perform hyperparameter tuning and feature selection simultaneously. In contrast, advanced techniques involve resampling methods like bootstrapping to evaluate feature stability and relevance across different data subsets. Resampling helps reduce variance in feature selection, making the process less susceptible to outliers or noise in the data. Moreover, joint optimization of feature selection and model parameters can potentially lead to improved model performance and interpretability, particularly in high-dimensional datasets with complex feature interactions. In this study, we compared some current bootstrapping-based methods with basic feature selection methods by assessing their prediction performance in identifying biomarkers related to human disease. This comparison aimed to identify the best method to deepen our understanding of complex biological systems and enhance disease diagnosis and treatment strategies.

3 - Diffusion Domain Adaptation from Source to Target via Schrödinger Bridge

Jinju Park, Ajou University, Suwon, Korea, Republic of, Hyunjung Shin

Labeling newly generated data has been attracted researchers' interests. One widely used approach is domain adaptation, which labels an unlabeled target domain from plentiful labels in the source domain by reducing the distribution discrepancy between the two domains. To narrow the distribution discrepancy, we can leverage one of the most recent generative models, the diffusion model. The diffusion model learns to converse from a data distribution to a Gaussian distribution, and in the reverse process generates data from a Gaussian distribution to a data distribution. Domain adaptation can be achieved naturally by assuming the data distribution in the conversion process as the source domain and the data distribution in the reverse process as the target domain. However, there is a computational burden incurred during the conversion to Gaussian and the reverse process from Gaussian. To avoid the difficulty, we propose a domain adaptation method that employs a Schrödinger bridges for the diffusion model. The proposed method bypasses the Gaussian distribution and directly converts the data distribution in the source domain to the data distribution in the target domain. Experiments show that the proposed method can significantly improve the performance of target domain by effectively adapting the data distribution of source domain.

4 - Towards Improving the Interpretability of Deep Learning

Zhen Lin, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, Rama Ramakrishnan

We develop low depth (up to depth 8) classification trees with hyperplanes to closely approximate neural networks. In this way, we contribute in increasing the interpretability of neural networks. To facilitate the stronger performance of the trees, we develop an optimization-based adaptive learning method to generate additional data for training the trees. We also propose variants of our method based on randomization and the combination of optimization and randomization. This method improves the performance of the trees in approximating the neural networks, and in real-world classification tasks. We show this optimization method performs better than randomization. We report computational results on 59 real-world classification datasets, with different sizes of neural networks with accuracy $\sim 95\%$ relative to the neural network.

ME69

Regency - 607

Technology and Business

Invited Session

eBusiness

Chair: Xiaoning Wang, University of Texas at Dallas, 851 Greenside Dr., Richardson, TX, 75080, United States

1 - Breaking Terrains or Scaling Heights: The Role of AI in Flat versus Hierarchical Startups

Xiaoning Wang, University of Texas at Dallas, Richardson, TX, United States

We examine whether artificial intelligence (AI) facilitates the success of flat startups or hierarchical startups. We explore the interaction effect between AI adoption and startup hierarchy by combining startup data from Crunchbase, patent application data from USPTO, and detailed employee job descriptions from LinkedIn. We find that while the adoption of AI helps flat companies achieve better commercial success, the adoption of AI also helps hierarchical companies achieve better innovation performance. The mechanism tests indicate that, on the one hand, AI can improve the organization's coordination capabilities by automating specific sections in the operation and production process, thereby streamlining the company's operational management, avoiding the disorder of flat organizations, and better achieving economies of scale. On the other hand, using AI can also enable companies to expand the scope of information acquisition and integration, thereby breaking the barriers of information sharing within hierarchical companies and improving innovation performance. These results not only provide new evidence for the interaction between AI and organizational structures but also demonstrate that AI, as a general-purpose technology, has multiple application formats and corresponds to a variety of management models depending on the types of technology used and the characteristics of the establishment within each firm.

2 - Exploring the Dynamics of Hashtag Use and Comment Toxicity on Social Media

Anna Gao, University of Pennsylvania, Philadelphia, PA, United States

The advent of social media has revolutionized how information is disseminated and discussed globally. This project explores the role of informational cues, such as hashtags used by content creators, in fostering safe online communities, with a particular focus on comment toxicity in social media discussions. Using Twitter's API, we collected tweets and engagement metrics from a week-long sample, and used the Trump Trial as an initial case study. We aim to understand what drives toxic comments during ongoing, high-profile events. In particular, we apply zero-shot learning to classify tweets into categories of support, opposition, or neutrality towards Trump and discover that tweets opposing Trump receive higher engagement and more negative sentiment. Additionally, we assess hashtag relevancy using a pre-trained language model and find that higher relevancy scores correlate with increased tweet impressions, while greater variance in relevancy scores reduces engagement. Our preliminary analysis reveals key insights into hashtag relevancy and comment toxicity. We found a positive correlation between average relevancy and average comment toxicity, indicating that less relevant hashtags to the tweet's content are less likely to attract toxic comments. These findings suggest that the strategic use of relevant hashtags can mitigate toxic interactions and foster healthier online discussions. By understanding and leveraging these dynamics, content creators can make smarter decisions to create a safer digital world.

3 - Data and Algorithm: Designing Marketplace Analytics for Platform Sellers

Yi Liu, University of Wisconsin-Madison, Madison, WI, United States, Fei Long

The rapid growth of e-commerce giants and the abundance of data has spurred the development of AI-powered analytics services, such as competitive intelligence and automated pricing, which enables marketplace sellers to make informed, data-driven decisions. Third-party providers (e.g., Jungle Scout and Helium 10) compete with platforms themselves (such as Amazon's brand analytics) in offering marketplace analytics. Yet we are witnessing platforms adopt various strategies in sharing data with third-party analytics providers, ranging from restrictive to permissive (e.g., permitting data scraping) with some even actively facilitating (e.g., API sharing). In this paper, we ask *why* and *how* an e-commerce platform may benefit from sharing data with third-party providers, despite the platform's inherent advantages in data access, and the capability to design its own analytics services for better control over sellers' actions. We also study how platforms should design their analytics services, modeled as an algorithm to predict market competition levels. We find that the platform maximizes its own sales commissions by providing an over-optimistic analytics service. When market competition is moderate, this may lead to sellers' reluctance to use the platform's analytics service, resulting in a lose-lose situation, and prompting the platform to share data with third-party providers. However, in highly competitive markets, over-optimistic predictions can actually benefit sellers. In this case, the platform adopts a more restrictive data-sharing approach. Furthermore, when the platform can control the accuracy of third-party analytics via API access, it shares data in a larger parameter space, and it shares broader data when market competition intensifies.

4 - Memory and Beliefs in Financial Markets: A Machine Learning Approach

Zhongtian Chen, University of Pennsylvania, Philadelphia, PA, United States, Jiyuan Huang

This paper shows that memory shapes investor beliefs and influences trading activities in financial markets. Using earnings forecasts by financial analysts as proxies for investor beliefs, we extract their memory and recalls by training a machine learning memory model. The extracted recalls strongly predict stock returns. Additionally, dispersion in recalled episodes predicts investor disagreement and trading volume, indicating that memory can be a powerful microfoundation for disagreement. We further document new facts about memory that are directly elicited from market data. Compared to an optimal benchmark trained to fit realized earnings, investor memory is distorted in two key ways. First, investors over-recall distant episodes in regular times but under-recall them during crisis times. Second, investor memory overweights the importance of past earnings forecasts, while the optimal benchmark mainly considers other firm fundamentals beyond earnings. These distortions in memory predict both forecast errors and key behavioral biases, including the presence of overreaction and underreaction.

ME70

Regency - 701

Advances in ACOPF

Invited Session

ENRE: Electricity

Chair: Daniel Bienstock, Columbia University, New York, NY, United States

1 - Scalable Multi-Period Ac Optimal Power Flow Utilizing Gpus

Mihai Anitescu, Argonne National Laboratory, Lemont, IL, United States, Sungho Shin, Vishwas Rao, Michel Schanen, D. Adrian Maldonado

There has been a growing interest in solving multi-period AC OPF problems, as the increasingly fluctuating electricity market requires operation planning over multiple periods. These problems, formerly deemed intractable, are now becoming technologically feasible to solve thanks to the advent of high-memory GPU hardware and accelerated NLP tools. This study evaluates the capability of the ExaModels.jl and MadNLP.jl tools for GPU-centered nonlinear programming to tackle previously unsolvable multi-period AC OPF instances. Our numerical experiments, run on an NVIDIA GH200, demonstrate that we can solve a multi-period OPF instance with more than 10 million variables up to 10−4 precision in less than 10 minutes.

2 - Tbd

Matias Villagra, Columbia University, New York City, NY, United States

tbd

3 - Alternating Methods for Decomposing Large-Scale AC OPF with Unit Commitment Decisions

Matthew Brun, Operations Research Center, MIT, Cambridge, MA, United States, Thomas Lee, Dirk Lauinger, Xin Chen, Andy Sun

In this work, we present an alternating penalty method tailored for solving large-scale, multi-period AC OPF problems with discrete unit commitment (UC) decisions. This method alternates between solving AC and UC problems, allowing for decomposition into a set of continuous, single period AC OPF problems and mixed-integer linear device level UC problems. We introduce a variety of additional techniques, including SOC relaxations, power ramping approximations, and line switching heuristics, that improve solution quality. We evaluate the effectiveness of this algorithm on test cases from the Grid Optimization Competition, Challenge 3, demonstrating competitive performance relative to other algorithm, and compare our solutions to a global lower bound.

ME71

Regency - 702

Learning and Optimization on Cyber-physical Systems

Invited Session

Data Mining

Chair: Wenbo Sun, University of Michigan Transportation Research Institute, Ann Arbor, MI, United States

Co-Chair: Yinan Wang, Rensselaer Polytechnic Institute, Troy, United States

1 - Fast Convergence of Retraction-Free Distributed Riemannian Gradient Descent

Youbang Sun, Northeastern University, Boston, MA, United States

We investigate non-convex optimization algorithms with orthogonal constraints. Conventional algorithms for this setting require either manifold retractions or other types of projection to ensure feasibility, both of which involve costly linear algebra operations (e.g., SVD or matrix inversion). On the other hand, infeasible methods are able to provide on par performance with higher computational efficiency. We propose a Decentralized Retraction-Free Gradient Tracking (DRFGT) algorithm, for which we establish both global and local convergence results over networks. Numerical experiments demonstrate that DRFGT performs on par with the state-of-the-art retraction based methods with substantially reduced computational overhead.

2 - Robust Continual Graph Completion for Adrd Knowledge Discovery

Shuteng Niu, Bowling Green State University, Bowling Green, OH, United States

Alzheimer's disease and related dementias (ADRD) are s group of neurodegenerative disorders that affect cognition, behavior, functional ability, and memory, with significant economic, medical, and social impacts. Despite extensive research, effective treatments and our understanding of ADRD remain insufficient. Recent advancements in non-pharmaceutical and pharmaceutical interventions, supported by modern computational methods, have shown promising results. Among state-of-the-art approaches, Knowledge Graph Completion (KGC) on graph-structured data demonstrates high potential in handling the complexity and multifunctionality natures of ADRD. Researchers have used the MEDLINE database to develop KG datasets and applied various KGC models, including traditional translational methods and advanced Graph Neural Networks (GNNs). However, current methods face two critical challenges, poor adaptation to growing KGs and vulnerability to adversarial attacks. Retraining larger KGC models is often prohibitively expensive, and graph-structured data is highly sensitive to adversarial attacks, leading to significant performance degradation in ADRD-related tasks, including but not limited to drug development, treatment planning, and risk prediction. Our research addresses these issues by developing a graph-based continual knowledge discovery framework for ADRD with robustness to adversarial attacks. With dynamic adaptation to growing KGs and resistance to adversarial attacks, our framework aims to improve the reliability and effectiveness of ADRD-related applications. This approach promises significant advancements in the field, providing more accurate and resilient tools to tackle this critical public health issue.

3 - Human Body Modeling for Vehicle Optimization via Inducing Points

Chang Li, University of Michigan, Ann Arbor, MI, United States, Wenbo Sun

Computer simulation models provide efficient and cost-effective alternatives for conducting physical experiments. For example, virtual crash testing simulations using morphed human body models (MHBMs) are well-adopted as anthropomorphic test devices are limited. However, due to computational constraints, it is infeasible to perform virtual crash testing across the entire population. This limitation demonstrates the need to select a subset of representative occupants for efficiently predicting population-wise injury risks. In this study, we introduce a novel method for selecting inducing points, which aids in determining the optimal size and placement of representative occupants. We employ the K-means clustering algorithm to identify inducing points that explain the variability within each gender's population. Additionally, we have developed Gaussian process (GP) models, based on these inducing points, for predicting population injury risks. We conducted a case study using MHBMs to analyze various injury outcomes under differing crash conditions. The results demonstrated that the selected representative occupants, optimized for size, accurately represent the entire population. Furthermore, these findings highlight the potential for efficiently achieving safety equity based on the model's predictions.

ME72

Regency - 703

Ethics in Action: Operationalizing Fairness in Algorithmic Decision-Making

Invited Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Finding Representative Group Fairness Metrics Using Correlation Estimations

Samuele Scaglioni, University of Illinois Chicago, Canoga Park, CA, United States, Hadis Anahideh, Nazanin Nezami

It is of critical importance to be aware of the historical discrimination embedded in the data and to consider a fairness measure to reduce bias throughout the predictive modeling pipeline. Given various notions of fairness defined in the literature, investigating the correlation and interaction among metrics is vital for addressing unfairness. Practitioners and data scientists should be able to comprehend each metric and examine their impact on one another given the context, use case, and regulations. Exploring the combinatorial space of different metrics for such examination is burdensome. To alleviate the burden of selecting fairness notions for consideration, we propose a framework that estimates the correlation among fairness notions. Our framework consequently identifies a set of diverse and semantically distinct metrics as representative of a given context. We propose a Monte Carlo sampling technique for computing the correlations between fairness metrics by indirect and efficient perturbation in the model space. Using the estimated correlations, we then find a subset of representative metrics. The paper proposes a generic method that can generalize to any arbitrary set of fairness metrics. We showcase the validity of the proposal using comprehensive experiments on real-world benchmark datasets.

2 - Individual-Level Explanation in the Context of Model Multiplicity

Parian Haghighat, University of Illinois at Chicago, Chicago, IL, United States, Hadis Anahideh

The phenomenon of Model Multiplicity (MM), also known as predictive multiplicity or the Rashomon effect, arises when multiple machine learning models perform equally well on a prediction task. Studies indicate that such models may exhibit significant internal differences, leading to inconsistent predictions for individuals. This poses challenges in providing counterfactual explanations (CEs), which are crucial for offering recourse recommendations. To address this issue, we propose a model-agnostic, individual-level solution involving the construction of uncertainty sets for optimal models. Our framework aims to select the best model for each individual while considering factors such as robustness of CEs and fairness, crucial for decision-making in high-stakes domains. These uncertainty sets are designed to encompass the most plausible explanations of the optimal model with high confidence. Theoretical and experimental analyses demonstrate that our framework satisfies properties lacking in existing methods, with minimal trade-offs in accuracy. We illustrate the effectiveness of our approach using real-world datasets, showcasing how our framework can enhance trust in model explanations.

3 - What Makes Systemic Discrimination, "Systemic?" Exposing the Amplifiers of Inequity

David McMillon, Emory University, Atlanta, GA, United States

Drawing on work spanning economics, public health, education, sociology, and law, I formalize theoretically what makes systemic discrimination “systemic.” Injustices do not occur in isolation, but within a complex system of interdependent factors; and their effects may amplify as a consequence. I develop a taxonomy of these amplification mechanisms, connecting them to well-understood concepts in economics that are precise, testable and policy-oriented. This framework reveals that these amplification mechanisms can either be directly disrupted, or exploited to amplify the effects of equity-focused interventions instead. In other words, it shows how to use the mechanics of systemic discrimination against itself. Real-world examples discussed include, but are not limited to, reparations for slavery and Jim Crow, vouchers or place-based neighborhood interventions, police shootings, affirmative action, and Covid-19.

4 - Finding Representative Group Fairness Metrics Using Correlation Estimations

Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

It is of critical importance to be aware of the historical discrimination embedded in the data and to consider a fairness measure to reduce bias throughout the predictive modeling pipeline. Given various notions of fairness defined in the literature, investigating the correlation and interaction among metrics is vital for addressing unfairness. Practitioners and data scientists should be able to comprehend each metric and examine their impact on one another given the context, use case, and regulations. Exploring the combinatorial space of different metrics for such examination is burdensome. To alleviate the burden of selecting fairness notions for consideration, we propose a framework that estimates the correlation among fairness notions. Our framework consequently identifies a set of diverse and semantically distinct metrics as representative of a given context. We propose a Monte Carlo sampling technique for computing the correlations between fairness metrics by indirect and efficient perturbation in the model space. Using the estimated correlations, we then find a subset of representative metrics. The paper proposes a generic method that can generalize to any arbitrary set of fairness metrics. We showcase the validity of the proposal using comprehensive experiments on real-world benchmark datasets.

5 - From Treatment to Outcome: Advanced Causal Inference for Resource Evaluation through Mediator Analysis and Counterfactual Estimation

Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

This paper introduces a novel framework that integrates causal inference and machine learning to enhance the analysis of treatment effects and resource allocation. The proposed approach employs a multistage methodology to bridge the gap between predictive modeling and causal analysis, addressing the challenges posed by the absence of counterfactuals in historical data. The framework consists of two primary stages: the first involves using machine learning models to establish pre-treatment baselines and identify at-risk observations, while the second applies causal inference techniques to evaluate the impact of interventions and assess post-treatment outcomes with machine learning. This integrated approach allows for a nuanced understanding of interventions' immediate and long-term effects, providing policymakers with actionable insights for more effective decision-making. The paper illustrates the efficacy of this framework through detailed case studies, including synthetic data and an educational setting focused on financial aid interventions. The results demonstrate the framework's potential to improve intervention strategies and resource allocation by offering dynamic, time-sensitive evaluations. The paper concludes with a discussion of the significance of the framework and recommendations for future research.

ME73

Regency - 704

Advanced Data Analytics for Energy Transition

Invited Session

Data Mining

Chair: Kailai Wang, University of Houston, Houston, United States

1 - Uncovering the Drivers of Alternative Fuel Vehicle Adoption in California: A Study of Latent Attitudes, Socio-Demographics, and Neighborhood Influences

Kailai Wang, University of Houston, Houston, TX, United States

Promoting the use of alternative fuel vehicles (AFVs) has become a long-term transportation strategy in California, which can bring a broad range of social, economic, and environmental benefits. Based on a sample of 3260 California residents from the 2018 California Panel Survey, this study explores the impacts of latent attitudes, socio-demographic characteristics, and neighborhood effects on consumers' current vehicle fuel type choice and their interest in purchasing or leasing an AFV in the future. One joint integrated choice and latent variable (ICLV) model is estimated to understand the taste heterogeneity within different population segments. The results suggest that latent attitudes towards environment, new technologies, car-utilitarianism, and residential location preference play critical roles in individuals' adopting new vehicle technologies. A range of socio-demographics, including age, race, gender, student status, education level, income level, household size, housing tenure, housing type and residential parking also make effects. Exposure to BEVs in both residential location and worksite has positive influence on AFV adoption, although public EV charging stations were not found to be essential factors since our respondents may mainly rely on home chargers. Moreover, the study suggests that individual's current user experience with AFVs has positive effect on their future interest in AFV. Overall, we predict the maximum market penetration of AFVs could be 41% of adult population. The findings offer detailed guidance on crafting California's transport policy to promote AFV, regarding the heterogeneity of the population's preferences and attitudes.

2 - Energy Equity-Aware Planning of Community Microgrid

Behnam Sabzi, University of Houston, Houston, TX, United States, Gino Lim, Jian Shi

As global energy systems face increasing challenges from climate change, social disparities, and infrastructure vulnerabilities, community microgrids have emerged as a critical solution for promoting sustainable and equitable energy access. This paper addresses the significant gap in energy equity within low-income neighborhoods through the strategic implementation of community microgrids, proposing three new community-based equity metrics. We introduce a two-stage stochastic mixed-integer programming model that effectively integrates the Community Equity Index (CEI) into the microgrid planning process. The first stage focuses on investment decision-making, while the operational planning subproblem is addressed in the subsequent stage utilizing Benders decomposition.

Our framework accounts not only for financial sustainability but also enhances the resiliency and environmental metrics of the target communities. The simulation results are promising, showing that our approach significantly outperforms traditional grid dependency by improving financial returns, reducing environmental impacts, and enhancing the resiliency of energy systems against uncertainties. Specifically, the proposed microgrid framework demonstrates a marked improvement in the Community Equity Index, ensuring that the benefits of energy advancements effectively reach underserved populations.

This study contributes to the literature by defining three new community-based equity metrics, providing a scalable and efficient solution technique for energy planning, and highlighting the role of equity considerations in community microgrid implementations. The findings are particularly relevant for policymakers and practitioners focused on integrating equity into energy planning and operations.

3 - Equitable Power Grid Restoration Considering Decision Maker's Risk Tolerance

Gino Lim, University of Houston, Houston, TX, United States, Behnam Sabzi, Jian Shi, saeedeh abbasi, Masoud Barati

This study addresses the urgent need for robust power system restoration strategies in the face of increasing environmental hazards such as hurricanes, floods, and tornadoes. As climate-related disruptions pose significant threats to power grid reliability, there is a critical demand for strategies that ensure equitable and efficient power restoration. This paper introduces a novel robust optimization (RO) model that integrates an equity-aware approach within a graph partitioning problem (GPP) framework, accommodating the varying risk tolerances of decision-makers concerning uncertain transmission line statuses.

4 - Power Demand Classification Using Pso-Based Ensemble Method

R. J. Kuo, National Taiwan University of Science and Technology, Taipei, Taiwan, Zhen-Xuan Xu

Time series data, where each data point is associated with time, finds applications in diverse fields such as engineering, management, and healthcare. The current study aims to propose a method to enhance the performance of time series classification. The proposed method utilizes MiniRocket to extract features from time series initially and then employs a particle swarm optimization (PSO) algorithm for an ensemble learning method with random forest (RF), XGBoost, and LightGBM for classification due to their explainability. The proposed method is evaluated using Italy's power demand data, comparing it with multi-layer perceptron (MLP), support vector machine, RF, XGBoost, and LightGBM in terms of accuracy, precision, recall, and F1-score. Additionally, various combinations of RF, XGBoost, and LightGBM with and without the PSO algorithm are also tested. According to the computation results, MLP achieves the highest accuracy for individual algorithms at 0.961089. For hybrid algorithms without the PSO algorithm, a combination of RF and LightGBM performs the best. However, when the PSO algorithm is combined, the hybrid of RF, XGBoost, and LightGBM outperforms all algorithms with an accuracy of 0.972095. Similar trends are observed for precision, where MLP outperforms individual algorithms, and the hybrid of RF and LightGBM performs well without the PSO algorithm. When combined with the PSO algorithm, the proposed algorithm, a hybrid of RF, XGBoost, and LightGBM, still outperforms all algorithms. The results for recall and F1 are consistent with those of accuracy and precision.

ME74

Regency - 705

Optimization and Analytics for Grid Reliability and Resilience

Invited Session

ENRE: Electricity

Chair: Feng Qiu, Argonne National Laboratory, Lemont, IL, United States

1 - State-of-Charge Estimation across Battery Chemistries: Insights from Unsupervised Domain Adaptation

Mohammad Badfar, Wayne State University, Detroit, MI, United States, Ratna Chinnam, Murat Yildirim

Accurate estimation of the state-of-charge (SOC) in lithium-ion batteries (LIBs) is paramount for the safe operation of battery management systems. Despite the effectiveness of existing SOC estimation methods, their generalization across different battery chemistries and operating conditions remains challenging. Current data-driven approaches necessitate extensive data collection for each battery chemistry and operating condition, leading to a costly and time-consuming process. Hence, there is a critical need to enhance the generalization and adaptability of SOC estimators. In this paper, we propose a novel SOC estimation method based on Regression-based Unsupervised Domain Adaptation. We evaluate the performance of this method in cross-battery and cross-temperature SOC estimation scenarios. Additionally, we conduct a comparative analysis with a widely-used classification-based unsupervised domain adaptation approach. Our findings demonstrate the superiority of the regression-based unsupervised domain adaptation method in achieving accurate SOC estimation for batteries.

2 - Prognostics approach for power asset maintenance

Shijia Zhao, Argonne National Laboratory, Lemont, IL, United States

Aging assets are observed across various power sectors. Almost half of the hydropower plants are exceeding 40 operational years. And we can easily find transformers and power lines exceeding 25 operational years. The changing operational patterns adds to the accelerated degradation. We find it is motivated or even urgent to propose a new generation of models to improve maintenance of power assets.

Periodic maintenance is easy to implement but lacks monitoring and inspection. Diagnostics can discover imminent failure risks but lack the ability to predict asset's failure time. That's why we are proposing prognostics to not only estimate degradation but also provide probability distribution of assets' failure time.

Based on the real-time sensor observations, we build degradation model to provide remaining lifetime distribution, which is used in dynamic maintenance cost model to provide just-in-time maintenance, Not too late, not too early. If we perform maintenance too frequently before actual failure, we waste O&M cost with premature maintenance. But if we perform maintenance too late and cause asset failure, we have to purchase costly replacement due to late maintenance. meaning with the estimates on failure time, we quantify the impacts of premature maintenance versus failure costs. Finally, the maintenance decisions are optimized at fleet level.

3 - Blackstart Power Distribution Grids using Distributed Energy Resources

Zhaoyu Wang, Iowa State University, Ames, IA, United States

This talk will introduce using inverter-based resources (IBRs) to blackstart power distribution grids. While transmission system blackstart has been well studied, power distribution system blackstart is a new topic that recently attracts industry interests thanks to the increasing penetration of IBRs. Because power distribution systems operate very differently than transmission systems, the existing blackstart work cannot be directly implemented. The blackstart of distribution systems requires coordination of grid-forming inverters, grid-following inverters, and intelligent switches, while considering unique characteristics of distribution grid operations, such as the radial topology, the three-phase unbalanced power flow, the stability, and the synchronization. We formulate this problem as a mixed-integer linear program which can be solved efficiently. We will show the performance on a modified IEEE 123-node test system. This talk will also briefly introduce our efforts to build Iowa's first renewable microgrid, which transforms an entire Iowa town to be a microgrid city with the best reliability and resilience.

4 - A generalized framework for resilient optimal sensor placement and fault diagnosis within energy systems

SARA KOHTZ, Binghamton University, Binghamton, NY, United States, Anabel Renteria

Accurate and efficient health monitoring for high power energy systems has become an imperative research area in the field of reliability engineering. Specifically, an optimal sensor network can enhance prognostics of engineering applications across many fields. These include but are not limited to electric vehicles, aircrafts, industrial manufacturing, and many other impactful concentrations. Therefore, determining optimal fault detection frameworks is an essential task. However, due to high cost and safety concerns, there is a scarcity of data from experimental testing and validations for energy systems. Thus, finite element simulation data is a necessity for determining the monitoring system. This study aims to develop a generalizable fault detection framework, which requires techniques that optimize signals processing and sensor placement on a fundamental level. In particular, the proposed method simultaneously determines optimal placement of sensors while training a predictor of faults. In addition, the case where a sensor fails is considered, ensuring one level of resilience for the chosen design. Various search algorithms are implemented to solve the generally applicable mathematical formulation, which utilizes predictor accuracy as the fitness function. The predictor can take on several forms for detection of faults, including a simple K-nearest neighbors, or a complex ensemble learning classifier. The strength of this proposed framework is demonstrated in engineering case studies, including fault detection on electric motors using hall-effect sensors, and partial discharge classification on transformers. Overall, this proposed method converges to a design that has high accuracy for detection of faults, and also satisfies a N-1 redundancy criterion.

ME75

Regency - 706

Advances in Data-driven Controls for Renewable Dominated Grids

Invited Session

ENRE: Electricity

Chair: Kyung-bin Kwon, Pacific Northwest National Laboratory (PNNL), Richland, United States

Co-Chair: Sayak Mukherjee, Pacific Northwest National Laboratory, Richland, WA, United States

Co-Chair: Hao Zhu, The University of Texas at Austin, Austin, TX, United States

1 - Risk-aware Structured Wide-area Control in Inverter-dominated Power System

Kyung-bin Kwon, Pacific Northwest National Laboratory, Richland, WA, United States

In this research, we develop a risk-aware controller for grid-forming inverters (GFMs) to minimize large frequency oscillations in GFM inverter-dominated power systems. To tackle the high variability from loads/renewables, we incorporate a mean-variance risk constraint into the classical linear quadratic regulator (LQR) formulation for this problem. The risk constraint aims to bound the time-averaged cost of state variability and thus can improve the worst-case performance for large disturbances. The resulting risk-constrained LQR problem is solved through the dual reformulation to a minimax problem, by using a reinforcement learning (RL) method termed as stochastic gradient-descent with max-oracle (SGDmax). In particular, the zero-order policy gradient (ZOPG) approach is used to simplify the gradient estimation using simulated system trajectories. Numerical tests conducted on the IEEE 68-bus system have validated the convergence of our proposed SGDmax for GFM model and corroborate the effectiveness of the risk constraint in improving the worst-case performance while reducing the variability of the overall control cost.

2 - Optimizing Distributed Energy Resources on a Data Budget

Vassilis Kekatos, Virginia Tech, Blacksburg, VA, United States

How much data are really needed to optimally schedule distributed energy resources (DERs)? To decide DER setpoints, a utility may have to solve an optimal power flow (OPF) every few minutes upon collecting values of uncontrollable load demands from hundreds or thousands of nodes. This work exploits redundancies in the data or features feeding the OPF as well as the structure of the OPF per se, to avoid communicating such data deluge and explore the trade-off between compression and grid's performance. The goal is to identify a small subset of critical features and use them to reconstruct a high-dimensional vector to feed the OPF. This task of OPF data selection and reconstruction is posed as a bilevel program wherein the outer level deals with data compression, and the inner level solves multiple instances of the OPF using the compressed data. To forego computationally cumbersome mixed-integer formulations, we put forth a group-Lasso scheme to effect feature selection and solve it via a proximal gradient scheme upon differentiating through the OPF mapping. Numerical tests demonstrate the computational advantage of the proposed scheme over mixed-integer alternatives and showcase the benefit of compressing data while having in mind the effect of compressed OPF data to the OPF solution.

3 - Feedback Optimization of Incentives for Distribution Grid Services

Andrey Bernstein, NREL, Golden, CO, United States

Energy prices and net power injection limitations regulate the operations in distribution grids and typically ensure that operational constraints are met. Nevertheless, unexpected or prolonged abnormal events could undermine the grid's functioning. During contingencies, customers could contribute effectively to sustaining the network by providing services. This talk proposes an incentive mechanism that promotes users' active participation by essentially altering the energy pricing rule. The incentives are modeled via a linear function whose parameters can be computed by the system operator (SO) by solving an optimization problem. Feedback-based optimization algorithms are then proposed to seek optimal incentives by leveraging measurements from the grid, even in the case when the SO does not have a full grid and customer information. Numerical simulations on a standard testbed validate the proposed approach.

4 - Stability-constrained Reinforcement Learning for Power System Control

Yuanyuan Shi, University of California San Diego, La Jolla, CA, United States

The past decade has witnessed success of learning and control methods such as reinforcement learning (RL) in a broad spectrum of applications. As a result, the application of RL in sustainability applications has attracted surging attention recently. Despite the promise, one of the biggest challenges for the deployment of RL in energy systems is the lack of **stability and performance guarantees**. Since energy systems are critical infrastructure, failure to maintain stability can lead to catastrophic consequences.

In this talk, I will present a series of our recent works on developing a stability constrained learning framework that combines policy learning in RL with Lyapunov stability in control to learn policy with formal stability guarantees. Specifically, I will highlight **monotonicity** as a key structure for both power system frequency control and voltage control. I will discuss how can we systematically construct the stability certificates via the Lyapunov approach, derive the structural constraints, and incorporate them into neural network design for guaranteeing stability in power system frequency and voltage control.

5 - Neural ODE-Model Predictive Control for Managing Renewables at the Grid Edge

Qiuhua Huang, Pacific Northwest National Laboratory, Richland, WA, United States, Patrick Salter

High penetration of distributed energy resources (DERs) in distribution systems can lead to fast voltage oscillations and violations issues. Conventional centralized voltage-var optimization and control cannot properly address these issues due to their relatively large control intervals and slow response of the control devices. This presentation introduces a distributed, AI-enabled fast voltage control method to mitigate fast voltage oscillations and violations caused by DERs by smoothing net load variations locally. The work leverages Neural Ordinary Differential Equations (NODE) for modeling residential building thermal dynamics and integrates this model with Model Predictive Control (MPC) within an end-to-end optimization framework to derive the optimal control policy for each residential house. Key findings include the NODE's capability to accurately track thermal dynamics and the success of the combined NODE-MPC in reducing net load variations by aligning HVAC operations with PV generation. Comparative analysis with deep reinforcement learning (DRL)-based centralized voltage control highlights the effectiveness of the NODE-MPC approach.

6 - Leveraging Predictions in Power System Frequency Control: An Adaptive Approach

Wenqi Cui, University of Washington, Seattle, WA, United States, Guanya Shi, Yuanyuan Shi, Baosen Zhang

Ensuring the frequency stability of electric grids with increasing renewable resources is a key problem in power system operations. In recent years, a number of advanced controllers have been designed to optimize frequency control. These controllers, however, almost always assume that the net load in the system remains constant over a sufficiently long time. Given the intermittent and uncertain nature of renewable resources, it is becoming important to explicitly consider net load that is time-varying.

This paper proposes an adaptive approach to frequency control in power systems with significant time-varying net load. We leverage the advances in short-term load forecasting, where the net load in the system can be accurately predicted using weather and other features. We integrate these predictions into the design of adaptive controllers, which can be seamlessly combined with most existing controllers including conventional droop control and emerging neural network-based controllers. We prove that the overall control architecture achieves frequency restoration decentralizedly. Case studies verify that the proposed method improves both transient and frequency-restoration performances compared to existing approaches.

ME76

Regency - 707

Power Systems Planning and Operation I: Integrating Multivariate Uncertainties

Invited Session

ENRE: Other Energy

Chair: Jordan Kern, North Carolina State University, Durham, NC, United States

1 - Novel Opportunities and Lessons Learned from Pnnl&Rsquo;S Grid Operations, Decarbonization, Energy and Environmental Equity Platform

Nathalie Voisin, Pacific Northwest National Lab., Seattle, WA, United States, The GODEEEP team

Meeting the Paris agreement’s ambitious goal to decarbonize the economy requires a paradigm shift in long-term planning across the entire energy sector. Bulk power grid long-term planning is traditionally performed in three steps: decarbonization roadmap design and scenario analytics, power plant siting and power grid resource adequacy studies, and finally reliability and economics studies. These steps are most often performed independently by different decision-makers (States, power system operators and regulators, and utilities), informed by inconsistent climate and socio-economics datasets across stakeholders. Consistent and accessible datasets and tools are critical to support the energy sector in their individual long-term planning towards overarching recommendations for resource adequacy, reliability, and cost efficiency. PNNL’s Grid Operations, Decarbonization, Environmental and Energy Equity Platform (GODEEEP) is a unique high-resolution, modular, and integrated modeling capability that evaluates the investment, operational, and equity challenges of alternative strategies to decarbonize the U.S. economy by 2050. This presentation will focus on the readily available tools, datasets and ways they can be leveraged through the lenses of end-use cases. This presentation will also provide the lessons learned and remaining gaps to support the communities in climate and equity informed energy plans and investments.

2 - Fairness in Post-Disaster Power Restoration

Benjamin Rachunok, North Carolina State University, Raleigh, NC, United States, Ignacio Sepulveda

The power grid Is essential to maintaining the basic functioning of a community. The frequency and intensity of extreme weather events has increased significantly worldwide and has been translated into longer, and more frequent power outages, causing social and economic damages. Evidence from previous restoration schedules has shown a significant disparity in the extent and duration of power outages experienced by low-income and minority groups, and rural communities. This reveals the necessity of including fairness into power restoration planning, translating equity policies into quantitative metrics and results. Current restoration strategies are based on experience, cost reduction, community resilience (time without power-based), or priority-based considering critical infrastructure (hospitals, fire stations, etc.), without considering customers’ characteristics. In this research, we address the power restoration planning problem for a transmission network considering customers’ vulnerability, captured by the CDC/ATSDR’s Social Vulnerability Index. In this problem, we dispatch a set of homogeneous repair crews to fix damaged components, in order to restore the power in an equitable manner considering operational constraints. The problem is modeled as a Mixed-Integer Linear Programming model, where equity is included in the objective function, testing different fairness perspectives. For each fairness perspective, the price of fairness is also computed in order to determine how expensive is the implementation of the corresponding strategy.

3 - Grace Foreseer: a Probabilistic Forecast Generator for Stochastic Unit Commitment

Dalia Patino-Echeverri, Duke University, Durham, NC, United States, Dimitrios Floros, Wei Wang, Mauricio Hernandez, Jordan Kern, Xiaodong Zhang

Increased probability of extreme weather events and ongoing and planned increases in the penetration of Variable Renewable Energy (VRE) resources in electric power systems highlight the need for improved uncertainty characterization and commitment and dispatch methods that account for such uncertainty.

We propose the GACE FORESEER which generates high quality Probabilistic Forecasts of weather variables, load, and variable renewable energy, for use in a Stochastic Unit Commitment (SUC). The outputs of the FORESEER are ensembles of time-series each with the same probability of occurrence. Each multivariable time series represents a plausible trajectory for weather variables, load, and wind solar energy production for each hour of a week-long planning horizon, in a specific geographical area.

Acknowledging that quality probabilistic forecasts for SUC cannot be presented as quantile trajectories -as often presented, with unrealistic temporal patterns of electricity demand (i.e., load) and/or renewable generation, where all hours have a load in the same percentile, the PFG forecasts are realistic as assessed by six metrics. These metrics focus on the Probability Distribution Function, the Autocorrelation Function, the geographic and variable cross correlations, the forecast errors, and the Continuous Ranked Probability Score.

4 - Impacts of Climatic and Socioeconomic Uncertainties on Net-Zero Power Systems Reliability: a Case Study from New York State

Vivek Srikrishnan, Cornell University, Ithaca, NY, United States

How do climatic and socioeconomic uncertainties impact the reliability of net-zero power systems? In this talk, I will provide an overview of recent research into how climatic, socioeconomic, and technological uncertainties jointly impact the performance of the planned future New York State (NYS) power system. NYS has one of the most ambitious electric power system decarbonization targets in the world, requiring net-zero emissions by 2040. To meet this goal, the New York Climate Action Council has created a scoping plan to expand the amount of renewable generation and storage to supplement hydropower from the Great Lakes. Stress-testing this plan against a variety of future climate and electrification scenarios provides insights into the reliability and vulnerability of decarbonizing and decarbonized power systems. I will highlight the impacts of these uncertainties on prices and load shedding, and share challenges posed by available climate data products for this stress testing.

5 - Balancing Accuracy and Computing Costs: Can We use Heuristics to Solve a Large-Scale Unit Commitment Problem?

Phumthep Bunnak, Cornell University, Ithaca, NY, United States, Stefano Coniglio, Stefano Galelli

Power system models have become increasingly larger and more complex, driven by use cases such as uncertainty analysis and multi-sectoral integration. These models are often run multiple times with different parameters, making runtime a bottleneck in the design of computational experiments. Currently, energy system modelers typically trade model accuracy with tractability by simplifying temporal, spatial, or technological dimensions. By leveraging the mathematical formulation of power system models, in this case a unit commitment problem, heuristics may offer an alternative pathway to reduce model runtime while maintaining an accurate system representation. Here, we explore this direction and contribute two novel computational frameworks to solve large-scale unit commitment problems: a rounding heuristic and a modified Dantzig-Wolfe decomposition. Computational experiments on country-scale grids reveal a speed-up reaching up to 10x while maintaining an optimality gap lower than 1%. The rounding heuristic proves to be highly scalable with problem size. Although runtime and solution quality may vary, these heuristics are not specific to any particular unit commitment problem.

6 - Deep Uncertainty and Deep Decarbonization: Scenario Discovery of Global Energy Transitions

Jacob Wessel, Tufts University, Medford, MA, United States, Gokul Iyer, Jonathan Lamontagne, Thomas Wild, Yang Ou, Haewon McJeon

Deep decarbonization is a primary objective for energy systems planners striving to meet national and global climate goals. Objectives, pledges, and obligations set forth by the Paris Agreement in both the short- (2030) and medium-term (2050-2060) pose a complex planning problem subject to deep uncertainty across multiple interconnected systems. Exploratory modeling analyses systematically mapping a wide outcome space through large scenario ensembles are needed to reveal robust strategies for decisionmakers to meet decarbonization goals. This study applies scenario discovery to 5,760 model realizations generated using a global multi-sector climate-economic model. The large ensemble systematically varies eleven energy-related uncertainties, including techno-economic parameters, institutional factors, and national mitigation pledges, among others. This broad collection of future states of the world is used to characterize the impacts of various uncertainties on national and global deep decarbonization pathways. Results show that energy transitions may be globally consistent while regionally disparate, as measured by multiple common energy system metrics. Larger economies and developing regions experience more severe economic outcomes, which can be exacerbated by regional investment risk throughout a wide uncertainty space. Further, the role of different CO_2 removal options in meeting mitigation pledges can vary widely across regions, though the global scale of CO_2 removal principally determines the budget for continued energy system emissions. Previous large-ensemble work has not (to the authors' knowledge) combined such a framework with national emissions pledges or institutional factors. This study highlights advantages of using scenario discovery for decision support as countries design strategies to meet future goals.

ME77

Regency - 708

Recent Advances in Stochastic and Robust Optimization

Invited Session

Computing Society

Chair: Haoming Shen, University of Arkansas, Fayetteville, AR, United States

1 - Two-Stage Distributionally Robust Optimization for Service Region Design in Crowdsourced Delivery

Aliaa Alnaggar, Toronto Metropolitan University, Toronto, ON, Canada, Merve Bodur

We consider a service region planning problem faced by a crowdsourced delivery platform where drivers are commuters who are willing to deviate from their original routes to a make a delivery in exchange for a compensation. The availability of drivers and service demand are uncertain, and it is difficult to estimate their probability distributions due to the absence of service history. To mitigate the effects of data ambiguity, we propose a two-stage distributionally robust optimization (DRO) model. The first stage selects which nodes to offer service in, while the second stage matches drivers and orders after uncertainty is realized. We derive an exact reformulation of the DRO model and develop a monolithic approximation based on a convex relaxation of the subproblem. We further strengthen the proposed approximation by a set of valid inequalities inspired by the linearization reformulation technique. The benefit of the proposed approach for service region design in crowdsourced delivery is demonstrated through numerical experiments.

2 - Residuals-Based Contextual Distributionally Robust Optimization with Decision-Dependent Uncertainty

Zoey Zhu, The Ohio State University, Columbus, OH, United States, Xian Yu, Guzin Bayraksan

3 - Value of Stochastic Solution with Right-Hand Side Uncertainty

Haoming Shen, University of Arkansas, Fayetteville, AR, United States

We revisit the value of stochastic solution (VSS) in the context of distributional ambiguity. When uncertainty arises from the right-hand side of a two-stage stochastic program, we investigate VSS's upper and lower bounds using distributionally robust and optimistic optimization. Because these bounds are NP-hard to compute exactly, we derive tractable yet tight relaxations for them and demonstrate their effectiveness through numerical experiments.

4 - Risk-Averse Reinforcement Learning for Economic Dispatch with Uncertain Demand Response

Bo Zhou, University of Michigan, Ann Arbor, MI, United States, Ruiwei Jiang, Siqian Shen, Xian Yu

We study risk-averse reinforcement learning (RL) for real-time economic dispatch with uncertain demand response (DR). The proposed risk-averse RL controls dynamic risk of a sequence of rewards in infinite-horizon Markov Decision Processes (MDPs), where we adapt the Expected Conditional Risk Measures (ECRMs) and prove time consistency. Using a combination of expectation and conditional value-at-risk (CVaR), we reformulate the risk-averse MDP as a risk-neutral counterpart with augmented action space and manipulation on the immediate rewards. We further prove that the Bellman operator is a contraction mapping, which guarantees the convergence of any value-based RL algorithms. Accordingly, we develop a risk-averse deep Q-learning framework to reduce variance and enhance solution robustness, and demonstrate our methods for real-time economic dispatch with uncertain DR on IEEE systems.

ME78

Regency - 709

Modern Results in Computational Statistics

Invited Session

Computing Society

Chair: Ilias Zadik, Yale, New Haven, CT, United States

1 - Deterministic Algorithms for Continuous Approximation Problems

Devin Smedira, Massachusetts Institute of Technology, Cambridge, MA, United States, David Gamarnik

We explore the applications of the correlation decay technique to design deterministic approximation algorithms for continuous problems with a bounded degree structure. We construct a quasi-polynomial time deterministic approximation algorithm for computing the volume of an independent set polytope with restrictions. Randomized polynomial time approximation algorithms for computing the volume of a convex body have been known now for several decades, but the corresponding deterministic counterparts are not available, and our algorithm is the first of this kind. The class of polytopes for which our algorithm applies arises as linear programming relaxation of the independent set problem with an additional restriction on each variable dependent on the maximal degree of the graph. We apply the correlation decay method to the natural discretization of this problem. We will also show how similar techniques can be applied to approximating the integral of classes of functions defined with respect to bounded degree graphical models.

2 - Power of two matrices of spectral algorithms for community detection

Souvik Dhara, Purdue University, West Lafayette, IN, United States

Spectral algorithms are some of the main tools in optimization and inference problems on graphs. In this talk, we discuss the power of spectral algorithms using two matrices in a simple graph partitioning problem with censored edges. It is shown that any spectral algorithm based on one matrix is suboptimal for most choices of the parameters, while spectral algorithms based on two matrices are optimal.

Based on joint work with Julia Gaudio, Elchanan Mossel, Colin Sandon

3 - The Low-Temperature MCMC threshold for sparse tensor PCA

Ilias Zadik, Yale, New Haven, CT, United States

Over the last years, there has been a significant amount of work studying the power of specific classes of computationally efficient estimators for multiple statistical models, including the class of low-degree polynomials, spectral methods and others. Despite that, our understanding of the important class in applications of MCMC methods remains quite poorly understood. For instance, for many models of interest the performance of even zero-temperature (greedy-like) MCMC methods is elusive.

I will describe you a recent result that that we fully characterize the performance of low-temperature MCMC methods for the case of the sparse tensor PCA model. In particular, we discover that it the required signal to noise ratio requires for these MCMC methods to work is significantly higher than the performance of the best known polynomial-time estimator.

4 - Smoothed Analysis for Learning Concepts with Low Intrinsic Dimension

Vasilis Kontonis, University of Texas at Austin, Austin, TX, United States, Gautam Chandrasekaran, Adam Klivans, Raghu Meka, Konstantinos Stavropoulos

In the well-studied agnostic model of learning, the goal of a learner-- given examples from an arbitrary joint distribution on $\mathbb{R}^d \times \{\pm 1\}$-- is to output a hypothesis that is competitive (to within ϵ) of the best fitting concept from some class. In order to escape strong hardness results for learning even simple concept classes in this model, we introduce a smoothed analysis framework where we require a learner to compete only with the best classifier that is robust to small random Gaussian perturbation.

This subtle change allows us to give a wide array of learning results for any concept that (1) depends on a low-dimensional subspace (aka multi-index model) and (2) has a bounded Gaussian surface area. This class includes functions of halfspaces and (low-dimensional) convex sets, cases that are only known to be learnable in non-smoothed settings with respect to highly structured distributions such as Gaussians.

Perhaps surprisingly, our analysis also yields new results for traditional non-smoothed frameworks such as learning with margin. In particular, we obtain the first algorithm for agnostically learning intersections of k-halfspaces in time $k^{\poly(\frac{\log k}{\epsilon \gamma}) }$ where γ is the margin parameter.
Before our work, the best-known runtime was exponential in k.

Monday, October 21, 5:25 PM - 6:15 PM

Summit - Ballroom 3

Omega Rho Keynote Steve Eppinger

Plenary/Keynote Session

Keynote

Chair: Jose Correa, Universidad de Chile, Santiago, Chile

1 - Adventures in Engineering Management

Steven Eppinger, Massachusetts Institute of Technology, Cambridge, MA, United States

There are many interesting problems in engineering management for which network modeling and management science methods can be usefully applied - particularly where complex systems are involved. In this Omega Rho Distinguished Lecture, I will describe some problems addressed using Design Structure Matrix methods, challenges of Agile development for complex systems, and more..

Summit - Ballroom 2

The Importance of Lived Experience for Better, Smarter OR Models to Disrupt Sex Trafficking

Plenary/Keynote Session

Keynote

Chair: Maria Mayorga, North Carolina State University, Raleigh

1 - The Importance of Lived Experience for Better, Smarter OR Models to Disrupt Sex Trafficking

Thomas Sharkey, Clemson University, Clemson, SC, United States

Given the complexity and range of experiences in human trafficking, research and practice in this space has the potential for unintended consequences. Even well-meaning research and action can cause great harms if based on fuzzy conceptualization, overgeneralizations, or incomplete data. This talk will demonstrate the importance of including experts with lived experience, commonly referred to as survivors, in creating OR approaches to disrupt sex trafficking. It will show how our conceptualization of sex trafficking became better (more accurate, attuned to context, and actionable) and how our OR models and analysis became smarter because of our inclusion of experts with lived experience as research partners. The talk will then discuss how the OR and analytics community can include people with lived experience in our community's efforts to help disrupt sex trafficking.

Summit - Ballroom 1

IFORS Distinguished Lecture

Plenary/Keynote Session

Keynote

Chair: Janny Leung, The University of Macau, Taipa, N/A, Macau

1 - From Lotteries to Optimal Decision Making: Operations Research and the Theory of Moments

Chung Piaw Teo, National University of Singapore Business School, Singapore, Singapore

Lottery games, with their complex probabilistic structures and significant financial stakes, provide a fertile playground for Operations Research (OR) researchers. This talk explores rich opportunities for applying advanced OR techniques to optimize various aspects of lottery operations, from ticket sales to payout management, and illustrates the broader implications for decision making in related fields. Through various real life case studies, we demonstrate how the study of lottery games can yield insights into human behavior, statistical anomalies, and effective risk management strategies. In particular, we show mean-variance analysis, a core concept within the theory of moments, can be a powerful method in decision making within this sector. This method evaluates the trade-off between risk and return, providing a structured and quantitative approach to optimize decisions under uncertainty.
We also develop a general framework for the “picking winners” problem, aiming to select a small pool of candidate solutions to maximize the chances that one will perform exceedingly well ina combinatorial optimization problem, under a linear and additive random payoff function. This problem is formulated using a two-stage distributionally robust model with a mixed 0-1 semidefinite program, leveraging the Theory of Moments. This approach allows us to exploit the diversification effect to improve the chances of achieving a high ex-post payoff.

Applications of this framework in fields such as revenue management (offer construction) and fraud detection (credit rating) will be presented. Additionally, we demonstrate how this framework is connected to the field of explainable AI, enhancing transparency and trust in the decision-making process.

Tuesday, October 22, 8:00 AM - 9:15 AM

TA01

Summit - 320

Recent Advances in Human – AI Interactions

Invited Session

Service Science

Chair: Agni Orfanoudaki, Oxford University, Oxford, OX26SQ, United Kingdom

1 - Algorithmic Assistance with Recommendation-Dependent Preferences

Jann Spiess, Stanford GSB, Stanford, CA, United States, Bryce Mclaughlin

When an algorithm provides risk assessments, we typically think of them as helpful inputs to human decisions, such as when risk scores are presented to judges or doctors. However, a decision-maker may not only react to the information provided by the algorithm. The decision-maker may also view the algorithmic recommendation as a default action, making it costly for them to deviate, such as when a judge is reluctant to overrule a high-risk assessment for a defendant or a doctor fears the consequences of deviating from recommended procedures. To address such unintended consequences of algorithmic assistance, we propose a principal-agent model of joint human-machine decision-making. Within this model, we consider the effect and design of algorithmic recommendations when they affect choices not just by shifting beliefs, but also by altering preferences. We motivate this assumption from institutional factors, such as a desire to avoid audits, as well as from well-established models in behavioral science that predict loss aversion relative to a reference point, which here is set by the algorithm. We show that recommendation-dependent preferences create inefficiencies where the decision-maker is overly responsive to the recommendation. As a potential remedy, we discuss algorithms that strategically withhold recommendations, and show how they can improve the quality of final decisions.

2 - Human-AI Interactions and Societal Pitfalls

Jian Gao, UCLA Anderson School of Management, Los Angeles, CA, United States, Francisco Castro, Sebastien Martin

When working with generative artificial intelligence (AI), users may see productivity gains, but the AI-generated content may not match their preferences exactly. To study this effect, we introduce a Bayesian framework in which heterogeneous users choose how much information to share with the AI, facing a trade-off between output fidelity and communication cost. We show that the interplay between these individual-level decisions and AI training may lead to societal challenges. Outputs may become more homogenized, especially when the AI is trained on AI-generated content. And any AI bias may become societal bias. A solution to the homogenization and bias issues is to improve human-AI interactions, enabling personalized outputs without sacrificing productivity.

3 - Automated Feature Extraction for the Sts National Database: the Impact of Artificial Intelligence

Periklis Savvas Petridis, Massachussetts Institute of Technology, Cambridge, MA, United States, George Margaritis, Agni Orfanoudaki, Dimitris Bertsimas

Cardiothoracic surgery departments and clinics in the US rely on central agencies, such as the Society of Thoracic Surgeons (STS), to evaluate their operational performance compared to their peer institutions. Specifically, 97% of U.S. adult cardiac surgery programs collaborate and transfer their patient data to the STS National Database for quality improvement and risk assessment. However, as patient records primarily consist of unstructured text reports, such data transfers involve a large operational overhead, requiring manual data extraction by teams of experienced data managers. Utilizing recent breakthroughs in Natural Language Processing, we propose an end-to-end machine learning pipeline that automatically extracts all patient data (either structured or unstructured) from multiple sources, over multiple patient visits, and for multiple target outcomes. Our pipeline can be extended to new sources or outcomes, as hospitals may use different data conventions from each other. Preliminary results on Massachusetts General Hospital data show promise and our methodology achieves up to 98% AUC in common diagnoses and up to 85-95% in more challenging ones. We believe using an automated pipeline for data extraction has four benefits: i) reduce operational overhead and costs for institutions when transferring their data, ii) increase data consistency and quality, while reducing variation from human errors, iii) allow institutions to transfer even richer data that would otherwise require additional human effort, and iv) create a unified framework for end-to-end medical predictions and diagnoses from any type and format of patient medical reports.

4 - Learning the Minimal Representation of a Continuous State Space Mdp from Transition Data

Omar Skali, QuantumBlack, Boston, MA, United States

We propose a novel framework for learning a concise MDP model of a continuous state space dynamic system from observed transition data. Our approach relies on partitioning the system's feature space into regions constituting states of a finite deterministic MDP representing the system. We introduce an in-sample property on partitions of the feature space we name coherence and show that if the class of possible partitions is of finite VC dimension, any coherent partition with the transition data converges to the minimal representation of the system with provable finite-sample PAC convergence guarantees. This theoretical insight motivates our Minimal Representation Learning (MRL) algorithm.

5 - Does Robot-Assisted Surgery improve healthcare delivery?

Sung Joo Kim, Purdue University, West Lafayette, IN, United States, Jianing Ding, Susan Lu

We investigate the role of an emerging AI-based clinical treatment method, robot-assisted surgery (RAS), in transforming the healthcare delivery. In this study, we first examine the effect of RAS on clinical outcomes. Further, investigating the RAS implementation, we show the synergy of RAS implementation in alleviating healthcare racial disparity.

TA02

Summit - 321

Navigating Career Pathways in Industry: Insights and Strategies

Panel Session

Quality, Statistics and Reliability

Co-Chair: Jia Liu, Auburn University, Auburn, United States

1 - Moderator Panelist

Jia Liu, Auburn University, Auburn, AL, United States

2 - Panelist

Hongyue Sun, University of Georgia, Athens, GA, United States

3 - Panelist

Yifei Yuan, Amazon, Pleasanton, WA, United States

4 - Panelist

Shan Ba, LinkedIn, Mountain View, CA, United States

5 - Panelist

Lijuan Xu, Google, Sunnyvale, CA, United States

TA03

Summit - 322

Domain Knowledge in Learning Models: Relevance in Big Data Era

Invited Session

Quality, Statistics and Reliability

Chair: Raghav Gnanasambandam, Virginia Tech, Blacksburg, VA, 24060, United States

Co-Chair: Bo Shen, New Jersey Institute of Technology, Newark, NJ, United States

1 - Ml & Domain Knowledge with Applications in Manufacturing

Mengfei Chen, Rutgers University, Piscataway, NJ, United States, Weihong Guo

While deep learning algorithms in additive manufacturing have demonstrated efficacy in analyzing complex, massive data, the deployment of deep neural networks necessitates substantial datasets for training, posing challenges in real manufacturing scenarios. The limited availability of defect data leads to imbalanced datasets, especially in small-batch or one-of-a-kind manufacturing scenarios, significantly limiting the performance of predictive models. While dealing with limited data from one type of defect is already non-trivial, the issue of imbalanced dataset becomes even more challenging when multiple defects are unbalanced (some defects occur more than others). To address these challenges, this study introduces a conditional DCGAN-based data augmentation method that ensures the physical validity of synthetic in-process sensor data. The model incorporates domain knowledge such as layer information, temperature distribution, and the shape of the melt pool. Physics constraints are integrated into the custom architecture of the generator, and penalty terms are added to the loss function to prevent the production of outputs that violate physics constraints. This approach not only reduces reliance on historical data but also enhances defect prediction accuracy.

2 - Sparsest: Exploiting Data Sparsity in Spatiotemporal Modeling and Prediction

Junfeng Wu, Rensselaer Polytechnic Institute, Troy, NY, United States, Yinan Wang

Spatiotemporal data mining has a wide range of applications in modeling and predicting various complex physical systems (CPS), i.e., transportation, manufacturing, healthcare, etc. Among all the proposed methods, the Convolutional Long-Short Term Memory (ConvLSTM) has proved to be generalizable and extendable in different applications. However, ConvLSTM and its variants are computationally expensive, which makes it inapplicable in edge devices with limited computational resources. With the emerging need for edge computing in CPS, efficient AI is essential to reduce the computational cost while preserving the model performance. Common methods of efficient AI are developed to reduce model redundancy (i.e., model pruning, etc.). However, model redundancy is limited in spatiotemporal data mining as the embedded dependencies are complex and hard to capture. Instead, there is a fair level of data redundancy, which has been largely overlooked in existing research. Using the sequence of images as an example, (1) the informative pixels at each frame are usually spatially sparse (i.e., with a fixed background), and (2) the informative features in the first-order difference of two adjacent images are possibly sparse due to the slow temporal evolvement. Therefore, we develop a novel efficient ConvLSTM that pioneering exploits the sparsity of informative features. More specifically, the sparse convolution and the Delta algorithm are incorporated to exploit the spatial sparsity at each single time step and the sparsity in the first-order difference over a sequence, respectively. The experiment demonstrates that our proposed methods can achieve comparable performance to the original ConvLSTM with reduced computational complexity.

3 - Sequential Optimization Based on a Networked System

Siqin Yang, Arizona State University, Tempe, AZ, United States, Hao Yan

Composite Bayesian optimization that leverages insights from the internal components of objectives has demonstrated excellent performance across various applications. Using directed acyclic graphs, these networked systems can be optimized sequentially. For instance, an additive manufacturing process can be mapped as a sequential printing process with output from last layer would serve as input for the future layer. This presentation will concentrate on sequential optimization of the process parameters for each layer, employing rollout as a value approximation to estimate the reward for each action. By utilizing internal components in our graph models, we propose a method for making dynamic decisions applicable to various scenarios.

4 - Physics-Guided Auto-Regressive Neural Network with eXogenous Inputs

Xinchao Liu, Georgia Institute of Technology, Atlanta, GA, United States, Yifeng Wang, Chuck Zhang, Xiao Liu

It is challenging to train a physics-guided neural network with time-varying exogenous inputs for transient dynamical systems. In this work, we propose a novel approach, Physics-Guided Auto-Regressive Neural Network with eXogenous Inputs, to learn the temporal evolution of system states controlled by time-varying exogenous inputs such as heat sources and pollution releases.

TA04

Summit - 323

Advanced Statistical Models and Decision Making

Contributed Session

Chair: Wenlan Xie, The University of Sydney, Sydney, N/A, Australia

1 - Quickest Change Detection in Multi-Stream Unnormalized and Score-Based Statistical Models

Wuxia chen, University of Pittsburgh, Pittsburgh, PA, United States, Sean Moushegian, Taposh Banerjee, Vahid Tarokh

This paper introduces an approach to mult-stream quickest change detection for unnormalized and score-based statistical models. Traditional optimal algorithms in the quickest change detection literature require explicit pre-change and post-change distributions to calculate the likelihood ratio of the observations, acquiring which could be computationally costly and sometimes even not feasible for complex machine learning models. To address these challenges, we propose the min-SCUSUM method, which is a Hyvarinen Score-based algorithm. The Hyvarinen Score can be calculated for unnormalized and score-based models. We provide delay and false alarm analysis of the proposed algorithm and show that the asymptotic performance of the algorithm depends on the Fisher divergence between the pre- and post-change distributions.

2 - Connections to Consumption: Fusing Network Analysis and Machine Learning for Premium Wine Market Forecasts

Bryan Boots, University of Missouri-Kansas City, Kansas City, MO, United States

This study develops a predictive model for estimating global demand for Chilean wine, utilizing a combination of classical machine learning methods, natural language processing (NLP), and network (graph) methods. The ensemble and nested models draw from diverse data sources, including a structured dataset of wine exports from Chile, unstructured text-based product reviews, insights derived from network/graph models that represent complex market interactions, and macroeconomic and demographic data.

The methodology integrates these heterogeneous data forms to understand factors related to global demand for Chilean wine. Machine learning algorithms predict sales volumes, while NLP techniques analyze sentiments, topics, and trends from customer reviews to gauge consumer preferences across national markets. Network methods elucidate the relationships between market actors and their impact on wine sales.

The model’s robustness is tested through against historical data. In an industry where long-term capital investments are crucial – newly-planted grapevines take at least three years to produce commercially viable grapes, and up to ten years for a new vineyard to reach breakeven – even small improvements in demand forecasting can significantly impact business outcomes. This model aids stakeholders in strategic planning and inventory management and offers insights for sales and marketing professionals in competitive B2C or B2B2C environments.

By advancing hybrid analytical methods in market demand forecasting, this work contributes to consumer analytics, demonstrating how integrated data-driven approaches can enhance the understanding of consumer behavior and market dynamics. This approach provides a novel contribution to both the academic literature and practical applications in market demand forecasting.

3 - Classification of Unbounded Data by Gaussian Mixture Models via deep ReLU Networks

Tian-Yi Zhou, Georgia Institute of Technology, Atlanta, GA, United States, Xiaoming Huo

This paper studies the binary classification of unbounded data from R^d generated under Gaussian Mixture Models (GMMs) using deep ReLU neural networks. We obtain — for the first time — non-asymptotic upper bounds and convergence rates of the excess risk (excess misclassification error) for the classification without restrictions on model parameters. While the majority of existing generalization analysis of classification algorithms relies on a bounded domain, we consider an unbounded domain by leveraging the analyticity and fast decay of Gaussian distributions. To facilitate our analysis, we give a novel approximation error bound for general analytic functions using ReLU networks, which may be of independent interest. Gaussian distributions can be adopted nicely to model data arising in applications, e.g., speeches, images, and texts; our results provide a theoretical verification of the observed efficiency of deep neural networks in practical classification problems.

4 - Balancing the Costs and Benefits of Resilience-based Decision Making

, Weimar Ardila, Alex Savachkin, Daniel Romero, Jose Navarro

Most resilience models often prioritize measuring resilience using static, deterministic optimization techniques. However, robust post-disruption responses in social-physical (SP) systems require a focus on both rapid recovery and proactive mitigation of economic loss.} This paper proposes a stochastic dynamic optimization model based on an infinite horizon Continuous-Time Markov Decision Process (CTMDP) to balance the intervention costs and reduce the total recovery time ensuing from a disruption of an SP system. We aim to provide a model that can facilitate its application to different disruption scenarios. Our state-space formulation of the recovery process uses discrete performance intervals, whereby actions and resulting rewards/costs are related to investment resources, which govern state transitions. We illustrate the model via a testbed based on the 2010 Northern Colombia Dique Canal breach. Our results show that the optimal policy reduced the recovery time and restoration investment by approximately 40% and 10%, respectively, compared to the efficiency of government interventions. The proposed model features dynamic control of recovery resources and considers the costs of resilience assurance. The model can inform policymakers of ways to improve system resilience using balanced disruption recovery strategies.

5 - Guiding Smarter Decisions: AI-Driven Framework for Predicting Movie Success and Analyzing Dissemination Patterns

Wenlan Xie, The University of Sydney, Sydney, Australia, Haowen Tian

This study introduces a robust analytical framework integrating three advanced computational models to predict and analyze movie success dynamics comprehensively. Utilizing a dataset of 2,400 top-ranked IMDb movies spanning 2007 to 2024, alongside 270,000 reviews obtained through web scraping, this research employs a pre-trained Transformer model to perform sentiment analysis on movie reviews. These sentiments are subsequently processed through a Susceptible-Infectious-Recovered (SIR) model, effectively forecasting the dissemination rates of films. The integration of the SIR model, which has been demonstrated to enhance predictive capabilities by 85%, plays a pivotal role in developing unique dissemination features and vital inputs for a subsequent multi-task learning (MTL) model.

The MTL model leverages these inputs to predict movie success and initial box office revenue with refined accuracy, benefitting from the joint learning process that integrates structured data with dissemination metrics. This comprehensive modelling approach offers profound insights into the interrelations between consumer sentiment, societal influence, and market performance and significantly advances the film industry's predictive analytics capabilities. The rigorous integration of these models emphasizes the value of multidimensional analysis. It sets a new standard for predictive accuracy in film industry analytics, enabling stakeholders to make more informed decisions regarding marketing strategies, content alignment, and financial planning. This framework exemplifies the transformative potential of machine learning in enhancing industry-specific decision-making and academic rigour in social sciences research.

Keywords: Transformer Models; Sentiment Analysis; Multi-Task Learning; Susceptible-Infectious-Recovered (SIR) Model; Market Dynamics; Decision Support;

TA05

Summit - 324

Models for Corporate Decisions

Contributed Session

Chair: Yuanzhe Ma, Columbia University, New York, NY, United States

1 - Speed is the Essence of Warfare: the Impact of Rivalry on New Product Launches After An Innovative Shock

Fernando Campayo-Sanchez, University of Alicante, San Vicente del Raspeig, Spain, Francisco J. Mas-Ruiz, Juan Luis Nicolau

In this paper, we examine how product marketing decisions following an innovation shock may be affected by rivalry. We conceptualize innovation shock as the launch of a new innovative product design whose sales exceed the firm's expectations. To explain the entrant's reaction (especially in terms of speed of entry), we use product-level competitive comparisons disclosed by the companies. First, our research focuses on the awareness-motivation-capability (AMC) paradigm, anchored in the competitive dynamics literature. With competitive comparisons (whether at the visual, mechanical, or equipment level), incumbents try to convince consumers that their products are better than the products that entrants intend to market. In this scenario, the first hypothesis argues that an entrant will introduce a product earlier in a post-shock market position when it receives a comparative attack from an incumbent. The second hypothesis posits that, in a post-shock market, firms are more likely to accelerate the commercialization of their products when they develop an intentional cannibalization strategy (by disclosing competitive comparisons between their own products). To conduct the study, we analyzed the Spanish electric vehicle market between 2003 (the year in which the innovation shock occurred, coinciding with the commercialization of the second-generation Toyota Prius) and 2017. At the methodological level, we use mixed-effects Cox regression models and multinomial logit models with random coefficients to control for the existence of unobserved heterogeneity. Our results support the relationships assumed in the hypotheses and represent a breakthrough in the literature linked to competitive dynamics, social comparison, and intra-firm competition.

2 - Measuring and managing (un)sustainable workload under digitization: An activity analysis-based job demands-resources model

Ahmed-Youssef Oukassou, KU Leuven, Lille, France, Marijn Verschelde, Raluca Parvulescu

In this paper, we bridge the psychological Job Demands & Resources (JD-R) literature with the operations literature on activity analysis. The JD-R model provides a comprehensive framework, categorizing risk factors associated with different occupations. On one hand, resources positively affect performance, while on the other hand, demands relate to work conditions. Understanding the health impairment process (strain due to excessive demands) and the motivational process (resource-driven well-being enhancement) is essential in defining a sustainable level of workload. We apply a multidimensional nonparametric activity analysis on the JD-R model, to assess the relation between resources, demands, and mental well-being of employees. Using event-level data of railway control room operators, we construct a multidimensional JDR indicator for digital production, which we empirically validate by using proxies for operational risk and employee well-being.

3 - Does disruptive preference matter for team innovation? The conditional effects of compiled disruptive preference in scientific research teams

Xuansheng Wang, Renmin University of China, Beijing, China, People's Republic of, Qihan Yang, Zhiyang Liu, Jia Liu, Lifan Chen

This paper explores how team members' disruptive preference(DP) compiles to contribute to team innovation. Based on the interactional approach, we argue that team composition(compiled individual DP) affects innovation outcomes, and this effect differs between high-impact teams and low-impact teams. Using an open data set covering over 134M scientific publications, we conduct 2 studies to examine our hypotheses. Study 1 analyzes 184,811 research teams with papers in Nature, Science, PNAS or PLOS ONE published between 1900 and 2020, focusing on the effects of team mean DP on innovation outcomes and the moderating role of team influence. Study 2 examines 17,793 research teams that include Nobel Prize laureates, using polynomial regression and response surface analysis to investigate the effect of laureate-member DP congruence on innovation outcomes. The findings suggest: (1) team mean DP negatively affects outcome influence and positively affects outcome disruption. (2) Team influence positively moderates the relationship between mean DP and outcome influence, yet it negatively moderates the relationship between mean DP and outcome disruption. (3) In general, laureate-member DP congruence positively affects outcome influence and outcome disruption. (4) To high-impact teams, the congruence positively affects outcome influence; to low-impact teams, the congruence negatively affects outcome influence. This paper provides a novel perspective on team formation, emphasizing the role of compiled members' disruptive preference, and offers practical guidance for individuals with high innovative potential to assemble effective teams.

4 - The Distributional Effects of "Fulfilled by Amazon" (Fba)

Yuanzhe Ma, Columbia University, New York, NY, United States, Garud Iyengar, Thomas Rivera, Fahad Saleh, Jay Sethuraman

We provide an economic model of an e-commerce retail platform (e.g., Amazon) that offers a fulfillment service (e.g., "Fulfilled By Amazon") to merchants that sell on its platform. We demonstrate that the introduction of such a service generates disparate impacts on the different platform participants. In particular, while such a service benefits low service quality merchants and consumers, it reduces welfare for high service quality merchants. We compare the economic implications of the fulfillment service when it is introduced by the platform as opposed to when it is introduced by an independent logistics company. We show that the distributional welfare effects are more pronounced when the platform provides the fulfillment service thereby demonstrating the impact that dual ownership of the platform and fulfillment service has on participant welfare.

5 - Predicting Customer Lifetime using Machine Learning at Expedia Group

Alireza Inanlouganji, Expedia Group, Kirkland, WA, United States

Customer Lifetime Value (CLV) represents the customer's future cash flows over a long-term horizon, such as one year and beyond. Having the ability to estimate the future value of each customer enables businesses to make better decisions about customer acquisition, retention, incentives, marketing communications, long-term investments, and growth.

In this talk, we describe development, implementation and deployment of the CLV prediction models on the Unified Machine Learning platform at Expedia Group. The models are re-trained monthly and future value predictions for hundreds of millions of customers are updated daily and are consumed by business units and teams within Expedia Group.

TA06

Summit - 325

New Methods for Combinatorial Problems

Contributed Session

Chair: Raunak Sengupta, University of Illinois, Urbana Champaign, Urbana, IL, United States

1 - Estimating Optimal Solutions for the Knapsack Problem Through Regression and Machine Learning

Hugo Mainguy, University of Maryland, College Park, MD, United States, Bruce Golden, Luca Bertazzi

As with other NP-Hard problems, the knapsack problem can be computationally expensive to solve to optimality. Faced with this problem, we develop approaches to obtain estimated optimal solution values that are both very close to the optimal value and efficiently computable. Specifically, while leveraging multiple diverse datasets of various complexities, we find that, in line with previous work, we can obtain good accuracy using a linear regression model which includes the mean and standard deviation of simulations obtained using a randomized greedy heuristic as predictors. We can also include predictors based on the linear programming relaxation and a feasible solution of the integer program, and we can improve accuracy while maintaining reasonable computation times by applying random forest and neural network methods. The results are especially encouraging for more difficult datasets, where traditional estimation methods tend to produce disappointing results.

2 - Valid inequalities for cardinality constrained knapsack problems

Jinhak Kim, Ajou University, Suwon, Korea, Republic of

We study a continuous knapsack problem with an additional constraint that the number of nonzero components of the decision variable is no more than a fixed positive integer. This seemingly simple optimization problem is in fact NP-hard. When a branching scheme is applied to solve the problem to optimality, tight linear relaxations may reduce the number of enumeration and improve the computational efficiency. In this paper, we employ a logical disjunction that is equivalent to the cardinality constraint to develop valid inequalities. We first present general form of valid inequality implied by the disjunction. We next introduce some procedures that generate sequences of valid inequalities. Each generation procedure takes an existing valid inequality as input and returns another valid inequality when certain conditions are met. We also prove that the sequence of valid inequalities is finite. Then, we use the aforementioned disjunction to derive some valid inequalities known in the literature.

3 - Practical exact penalty and application to complexity analysis of penalty function methods

Shotaro Yagishita, The Institute of Statistical Mathematics, Tachikawa-shi, Japan, Masaru Ito

One of the typical approaches to constrained optimization problems is to use a penalty function. The theory of exact penalties allows the constrained minimization problem to be solved by a single unconstrained optimization problem. From a practical point of view, the implications of the existing theory are poor and not well connected to the theoretical analysis of the algorithm. We propose a practical theory of exact penalties. The theory allows for the analysis of the iterative complexity of penalty function methods of finding an approximate stationary point that satisfies the constraint.

4 - Beyond Traditional Robustness: Relaxing Connectivity for Robust Graph Partitions

Raunak Sengupta, University of Illinois, Urbana Champaign, Urbana, IL, United States, Rakesh Nagi, Ramavarapu S. Sreenivas

We address the challenge of generating robust solutions for the Balanced Connected k-Partition problem, where a connected graph is partitioned into k connected sub-graphs with equal node weights. This problem, arising in contexts like Area Coverage, Autonomous Construction, and Fair Allocation of Land, is known to be NP-Hard. We present provably good approximation algorithms that handle uncertainties in node weights. Although various robustness concepts exist in the literature, we demonstrate through pathological examples that these criteria can lead to undesirable solutions even when solved optimally. We establish two key structural results: 1) Robust solutions in practical settings require relaxing the connectivity constraint, and 2) A solution balancing loads and uncertainties simultaneously is highly robust and meets multiple robustness criteria within an acceptable approximation factor. Thus, finding a robust solution reduces to a two-dimensional load balancing problem on a graph with connectivity constraints, the two dimensions being node weight and uncertainty. For the case with 2 agents, we prove that a 1.5-approximation on both dimensions can be efficiently achieved and is the best possible guarantee. For the general case with m agents, we achieve a 2.66-approximation on both dimensions by relaxing connectivity constraints. Our approximation algorithms are supported by multiple interesting existential and structural results. Finally, an extensive numerical analysis is presented, where we demonstrate our algorithms' near-optimal performance and
ability to generate robust solutions.

TA07

Summit - 327

Advancements in Sensor and Data Applications

Contributed Session

Chair: Ziwei Liu, The UNIVERSITY OF TENNESSEE, Knoxville, TN, United States

1 - Electromyography signal synthesis for privacy protections using a generative adversarial network

Liwei Qing, North Carolina State University, Raleigh, NC, United States, Bingyi Su, Lu Lu, Sehee Jung, Xu Xu

Electromyography (EMG) signal analysis is essential for understanding muscle activity and fatigue, providing valuable insights into the prevention of work-related musculoskeletal disorders (WMSDs). However, the availability of open-access EMG databases is very limited because these data are collected from human subjects, raising privacy concerns regarding their sharing. To overcome this limitation, this study employs a generative adversarial network (GAN) to synthesize non-identified EMG signals for manual material handling (MMH) tasks, enhancing data availability for workplace safety and health studies while maintaining privacy. We analyze DoppelGANger, a prominent GAN variant for time series data and metadata generation, evaluating its efficiency and efficacy. The EMG dataset used for training and testing was sourced from our previous MMH experiment. To evaluate our synthetically generated dataset, we compare the statistical features of the real and synthetic datasets. The mean of the real data is 1.045×10^{-2} mv with a standard deviation of 1.553×10^{-2} mv , while the mean of the synthesized data is 1.084×10^{-2} mv with a standard deviation of 1.57×10^{-2} mv . The Wasserstein distance between the two datasets is 0.1755. These results demonstrate the effectiveness of using GAN models in generating high-fidelity EMG signals for MMH tasks, addressing the data shortage in occupational health research. Future research should incorporate more MMH task data and generate EMG signals conditioned on the input task type.

2 - Medical Machine Learning Model for Predicting 6-Min-Walk-Test Results

Sa Eun Park, Major in Industrial Data Science & Engineering, Department of Industrial Engineering, Pusan National University, Busan, Korea, Republic of, Ki Hun Kim, Myung Jun Shin, Tae Sung Park, Sang Hun Kim

The 6-minute walk test (6MWT) measures the walking distance of a subject during 6 minutes to evaluate the subject's exercise tolerance and functional aerobic capacity. It is the most established functional exercise capacity test among respiratory rehabilitation tests. The 6MWT results can be utilized to classify the subjects into one of normal gait capacity group (stable group) and low gait capacity group (risky group).

However, 6MWT requires not only the space for subjects to walk but also its examiners, which makes considerable costs. These costs will be significantly saved if only the persons expected to get risky results from 6MWT are pre-screened and then conducted for 6MWT. To realize the cost saving, this study develops a medical machine learning model for predicting the 6MWT results from the prior test results of 6MWT such as body composition test and physical function test. The model was designed to satisfy two essential requirements for its applicability to hospitals. Firstly, the inputs of the model were defined to be collectable from hospitals. Secondly, the model was interpreted to be understandable for clinicians to accept and utilize it. Specifically, this study identified key predictors of the model with Shapley values and derived an easily understandable function from the model via symbolic meta-modeling. The satisfaction of these two requirements makes the developed model operate well in real contexts of hospitals, which contributes to the establishment of pre-screening for 6MWT and enhance its cost efficiency.

3 - A Decision Support Framework for Building Energy Efficiency and Productivity Based on a Sensor Network

Ziwei Liu, University of Tennessee, Knoxville, Knoxville, TN, United States, Mingzhou Jin

This research focuses on optimizing energy consumption and improving occupant health and productivity in office buildings through the integration of a sensor network. Initially, a detailed building geometry model, inclusive of the HVAC system, is developed using OpenStudio. Subsequently, a sensor network is employed to gather real-world performance data from the building. This data is crucial for calibrating the developed model, ensuring that the simulations accurately reflect the building's actual energy consumption and performance characteristics. The study then conducts a series of experiments aimed at enhancing Indoor Air Quality (IAQ) and, by extension, occupant well-being and productivity. These experiments involve fine-tuning various parameters, including indoor air intake and occupant density, to find their impact on the indoor environment. Finally, the research seeks to find an equilibrium between minimizing energy usage and maximizing occupant health and productivity. This is achieved through the optimization of the HVAC system, focusing on temperature and lighting setpoints, air filtration, and fan power levels, alongside refining occupancy schedules. Overall, this study provides actionable insights and practical strategies in enhancing building efficiency while simultaneously ensuring occupant health and productivity are maintained at optimal levels.

TA08

Summit - 328

AI Innovations in Business and Security

Contributed Session

Chair: Farzaneh Moghadam, UNT, 1418 W Oak Street, Denton, TX, 76201, United States

1 - Artificial Intelligence in Cybersecurity: A New Paradigm Revolutionising Threat Intelligence and Defense Mechanism

Utkarsh Sinha, California State University, Long Beach, Long Beach, CA, United States, Laasya Priya Muktevi

The increasing sophistication and frequency of cyber-attacks necessitate a shift from traditional, reactive cybersecurity measures to more proactive and advanced defense mechanisms. This paper discusses the integration of Artificial Intelligence (AI) with Cyber Threat Intelligence (CTI), emphasising its transformative impact on cybersecurity. By automating the CTI cycle from requirement gathering to feedback. AI enhances the ability to detect cyber intrusions swiftly, automate threat responses, and refine defense strategies against emerging threats. We explore various AI techniques such as machine learning, deep learning, and natural language processing, which facilitate the identification of malicious patterns and activities with unprecedented speed and accuracy. Moreover, the paper highlights the need for secure implementation of AI technologies to prevent adversaries from exploiting AI-powered systems. Overall, the integration of AI with CTI represents a significant shift towards proactive cybersecurity, enabling organizations to not only anticipate but effectively counteract potential cyber threats.

2 - The Impact of AI on the Product to Market feedback loop in the technology industry.

Jiaqi Chen, Google, Chicago, IL, United States, Devika Naik, Nagesh Gulkotwar, Khyathi Balusu

The technology industry's swift evolution has seen AI become a transformative force in various aspects of product development and market analysis. This paper explores the transformative effects of AI on the product-to-market feedback loop, critical for market responsiveness and customer-centric product development.

Traditional product-to-market loops enable rapid product changes based on market insights, ensuring products are customer-focused. However, the process is hampered by time-consuming data collection, scalability issues, and potentially biased insights. These challenges hinder companies' ability to adapt to market shifts. AI can address these challenges by automating data analysis, enhancing scalability, minimizing bias and informing decision-making.

The paper utilizes a qualitative research approach, conducting online interviews with 20 participants across multiple process oriented roles in the US technology sector (e.g., product managers, analysts, customer success managers etc.). The literature review explores current AI methodologies which includes but is not limited to Generative AI, Natural Language Processing and Sentiment Analysis to offer research context. Data collection was conducted through Google Forms, facilitating the identification of trends related to the study's objectives.

Findings reveal that AI can enhance several stages of the feedback loop, including feedback capture, data retrieval, ticket management, data analysis, and generating recommendations. These AI-powered refinements streamline processes, improve accuracy, and enable faster responses to market shifts.

This research concludes that AI is set to transform the product-to-market loop in the technology sector. By empowering companies to innovate rapidly, anticipate customer needs, and thrive in a competitive landscape, AI is poised to redefine product development.

3 - Balancing Growth and Opportunity: AI's Role in Shaping Freelancers' Skills and Market Behavior

Eunah Cho, Baruch College, City University of New York, New York, NY, United States, Qiang Gao

Artificial intelligence (AI) has profoundly reshaped the freelance sector. This study investigates the intricate interplay between exploration and exploitation dynamics within the freelance realm in the context of ChatGPT adoption. Based on theories of exploitation and exploration, we first hypothesize how workers respond to the emerging ChatGPT in terms of skill development and how their responses affect their contracting and future performance. Using a comprehensive dataset from one of the largest online freelancer platforms, we employ a Difference-in-Differences approach to assess the impact of AI adoption. The study reveals a shift towards exploring skills among freelancers using ChatGPT, suggesting an adaptation to the automation of more routine tasks by AI. Concurrently, freelancers are experiencing an increased competitive advantage in bidding, as those effectively utilizing ChatGPT are enhancing their proposals or optimizing their bidding strategies to win more specialized tasks that align with their newly acquired skills. Despite these changes, work quality ratings have remained stable, indicating that the acquisition of new skills and an increase in winning bids have not compromised the quality of work delivered. However, a decline in time consumed to complete projects implies that freelancers can potentially boost their competitiveness in a market, showcasing their successful adaptation to technological advancements. This study contributes to our understanding of how AI technologies, exemplified by ChatGPT, are shaping the future of work in the freelance domain. This research also offers valuable insights for freelancers, platforms, policymakers, and other stakeholders navigating the evolving landscape of digital labor.

4 - Accelerating the Development of Digital Twins using Generative AI

Khaled Salah, Khalifa University, Abu Dhabi, United Arab Emirates, Senay Gebreab, Raja Jayaraman

Digital twins (DTWs), virtual representations mirroring physical assets, processes, or systems, have emerged as pivotal tools across many industries, aiding in predictive maintenance, optimization, and decision-making. However, their development often requires significant challenges, including data scarcity, model complexity, and computational demands. Generative AI (or GenAI) is a subset of AI techniques that focus on creating new data or content in the form of audio, video, images, or text. In this presentation, we demonstrate with examples and use cases how GenAI can accelerate the creation and development of DTWs. Specifically, we first show how GenAI can accelerate the generation of 3D models for DTWs by using 3D diffusion models to generate 3D assets from a single or few 2D images of an object. Second, we demonstrate how Large Language Model (LLM)-based assistants can automate the integration of data sources with a DTW and help establish communication between physical devices and the DTW model. Third, we illustrate how GenAI can facilitate code generation to streamline the development of data pipelines in DTWs by generating code for API integration, data processing, and storage. Fourth, we show how GenAI can enable interactive and dynamic visualization of DTW data and events based solely on natural language user input. Fifth, we test how generated data and synthetic scenarios can help in predictive simulation. Lastly, we highlight the use of GenAI in drawing actionable insights from collected past data to support decision-making processes.

5 - AI and Leadership

Farzaneh Moghadam, UNT, Denton, TX, United States

In industries across the globe, the integration of Artificial Intelligence (AI) is transforming leadership and management fundamentally. In this review, we explore the changing role of leaders in an AI-enhanced environment, highlighting the need for new competencies and ethical considerations. With AI technologies like machine learning and predictive analytics, decision-making is moving away from hierarchical models and toward data-driven frameworks, fostering a collaborative and innovative culture. Despite these challenges, however, transparent and accountable leadership is necessary to address algorithmic bias and privacy concerns. In today's workplace, AI is compulsory for employees, but what about leaders, CEOs, and managers?

This review synthesizes key findings from key studies, emphasizing the dual role AI can play in empowering organizations as well as replacing human leadership. Despite AI's benefits in strategic decision-making and operational efficiency, its limitations in replicating human emotional intelligence and ethical judgment underscore the need for human oversight.

Methodologically, our study collects comprehensive data on the intersection of AI and leadership by interviewing AI experts, business leaders, and HR professionals. An analysis of these interviews provides insight into organizational transformation, ethical challenges, and skills development. Leaders need to develop strategies that align AI capabilities with organizational goals, foster interdisciplinary collaboration, and embrace continuous learning.

In conclusion, navigating the AI-driven landscape demands a strategic vision, inclusive culture, and ethical accountability. Leaders who can effectively integrate AI into their decision-making processes while maintaining transparency and fostering innovation will be well-positioned to drive sustainable growth and maintain stakeholder trust in the AI era.

6 - Autonomous Agents and Monetary Policy: Can the Fed Run on Autopilot?

Emily Struble, University of Miami, Miami Herbert Business School, Coral Gables, FL, United States, Erotokritos Skordilis, Stefanos Delikouras

This work assesses the application of artificial intelligence (AI) on monetary policy. The goal of monetary policy, conducted by the Federal Reserve, is maximum employment, stable prices, and moderate long-term interest rates. One of the most important tools in monetary policy is the Fed interest rate, which is charged to banks on loans they receive from the Fed. The Fed rate is important because it can stimulate or contract the economy. For instance, during Covid and the 2007 economic crisis, the Fed rate was set to zero to foster economic growth. To the contrary, after Covid, the rate was increased to suppress inflation.

Reinforcement learning is a machine learning method that has recently gained traction in financial applications. Unlike other types of machine learning, reinforcement learning is a goal-directed method where the model (agent), through a continuous trial-and-error approach, learns to navigate its surrounding environment to maximize a cumulative long-term reward.

The scope of this work is to evaluate monetary policy decisions, namely the level of the Fed rate, suggested by scalable reinforcement learning agents, and compare them to the actual human-induced policies. This allows for performance assessment of machine learning in conducting monetary policy.

TA09

Summit - 329

Equity and Fairness in Online Decision Making

Invited Session

Auctions and Market Design

Chair: Will Ma, Columbia Business School, New York, United States

Co-Chair: Eric Balkanski, Columbia University, New York, NY, United States

Co-Chair: Andreas Maggiori, Columbia University, New York, NY, 10027, United States

1 - Fair Algorithms with Unfair Predictions

Andreas Maggiori, Columbia University, New York, NY, United States, Eric Balkanski, Will Ma

Algorithms with predictions, also called learning-augmented algorithms, is a recent framework for decision-making that leverages the power of machine-learned predictions without making any assumption about their quality. The goal is to improve the algorithm's performance when predictions are accurate while maintaining acceptable guarantees when predictions are erroneous. Due to challenges with fairness in machine learning, a serious concern is that learning-augmented algorithms could leverage biased predictions and, as a result, make unfair decisions. We study the classical secretary problem in the learning-augmented setting, and prove that such a concern is indeed legitimate. Current learning-augmented algorithms for that problem are initially given predictions about the candidates' values and select a candidate that is guaranteed to have value that is a $\max \{ \Theta (1) , 1 - O(\epsilon) \}$ approximation to the maximum value, where ϵ is the approximation error. However, these algorithms may pick the best candidate with vanishing probability. Our main result is a novel algorithm that achieves a $\max \{ \Theta (1) , 1 - O(\epsilon) \}$ approximation while also guaranteeing that the best candidate is selected with constant probability. We also extend our result to the k-secretary problem.

2 - Bridging Theory and Practice for Balancing Notions of Equity with Small Portfolios

Jai Moondra, Georgia Institute of Technology, Atlanta, GA, United States, Swati Gupta, Mohit Singh

In this work, we present a novel approach to address the challenge of choosing the right notion of fairness in various optimization problems. We introduce the concept of "portfolios" -- a set of solutions that contains an approximately optimal solution for each objective in a given class of objectives, such as Lp norms. This concept opens up new possibilities for getting around the “right” notion of fairness for many problems by summarizing the space of "good enough representative" solutions. We exemplify our approach for the fair facility location and scheduling problems. The former tries to balance access costs to open facilities that are borne by different groups of people, and the latter the load on machines. This can be done by minimizing an Lp norm or an ordered norm, however, there is no clear choice of “p” in the current literature, or the weights for an ordered norm. We demonstrate construction of small portfolios for these problems, and the trade-offs with the achievable approximation and the size of the portfolio. Further, motivated by the Justice40 Initiative that provides rolling budget investments, we impose a refinement-like structure on the portfolio for facilities opened over time. We develop novel approximation algorithms for these structured portfolios and show experimental evidence of their performance in US counties. We also present a planning tool that provides potential ways to expand access to U.S. healthcare facilities, which might be of independent interest to policymakers.

3 - Redesigning Service Level Agreements: Equity and Efficiency in City Government Operations

Zhi Liu, Cornell Tech, New York, NY, United States, Nikhil Garg

We consider government service allocation -- how the government allocates resources (e.g., maintenance of public infrastructure) over time. It is important to make these decisions efficiently and equitably -- though these desiderata may conflict. In particular, we consider the design of Service Level Agreements (SLA) in city government operations: promises that incidents such as potholes and fallen trees will be responded to within a certain time. We model the problem of designing a set of SLAs as an optimization problem with different equity and efficiency objectives under a queuing network framework; the city has two decision levers: how to allocate response budgets to different neighborhoods, and how to schedule responses to individual incidents. We: (1) Theoretically analyze a stylized model and find that the ``price of equity'' is small in realistic settings; (2) Develop a simulation-optimization framework to optimize policies in practice; (3) Apply our framework empirically using data from NYC, finding that: (a) status quo inspections are highly inefficient and inequitable compared to optimal ones, and (b) in practice, the equity-efficiency tradeoff is not substantial: generally, inefficient policies are inequitable, and vice versa.

4 -

Fair Ranking in the Cascade Model

Ozge Sahin, Johns Hopkins Carey Business School, Baltimore, MD, United States, Wentao Lu, Ruxian Wang

In the cascade model, the optimal solution is to rank sellers in decreasing order of revenues, which raises the concern of fairness as sellers with low-ranking positions receive little consumer attention. To alleviate this unfairness, we consider imposing different fairness constraints for the platform. We find that the problem after imposing minimum purchase probabilities on all sellers can be solved in polynomial time by solving the separation problem of the dual problem, which is equivalent to solving an optimal ranking for a cascade model with shifted revenues. We then consider the fairness constraints of imposing lower bounds of probability for each product to be placed in the top k positions. We identify conditions for the problem to be feasible. For the case of $k=1$, we show that the optimal solution is to offer n different rankings. For general case, we propose a simple heuristic with $1/2$ performance guarantee and a polynomial approximation scheme.

5 - Recurring Magician's Problem and Applications to Bayesian Task Assignment

Mohammad Reza Aminian, The University of Chicago, Booth School of Business, Chicago, IL, United States, Rad Niazadeh

Motivated by applications in online education platforms for curating and assigning tutoring sessions to teachers, we study a generalization of the online contention resolution scheme problem for uniform matroids—also known as the Magician's problem introduced in Alaei (2014). In this generalization, the decision maker sequentially interacts with elements that leave after each interaction and return at some point in the future. Once an element returns, the decision maker observes whether it is active or not and, if active, decides whether to select this element given a capacity constraint. Each time a selected active element returns, it also replenishes the capacity by one unit. The goal is to design an online algorithm that maximizes the probability that any element is selected each time the element returns, conditioned on the arriving element being active. We introduce a novel family of online algorithms that as an invariant induce a stationary distribution over the states. By studying this induced stationary distribution, we show that there exist simple polynomial-time online algorithms that obtain a minimum probability of $1 - \Theta(1/\sqrt{k})$ to select active elements each time they return or arrive for the first time, where k is the amount of capacity. As a corollary of our result, we also obtain a novel, yet simple and interpretable, asymptotically optimal online algorithm for the standard Magician problem in Alaei (2014), which might be of independent interest.

TA10

Summit - 330

Advanced Reinforcement Learning and Optimization Techniques

Contributed Session

Chair: Yilie Huang, Columbia University, New York, NY, United States

1 - Accelerating Low-Rank Factorization-Based Semidefinite Programming Algorithms on GPU

Qiushi Han, University of Illinois Urbana-Champaign, Urbana, IL, United States, Zhenwei Lin, Hanwen Liu, Caihua Chen, Deng Qi, Dongdong Ge, Yinyu Ye

In this paper, we address a long-standing challenge: how to achieve both efficiency and scalability in solving semidefinite programming problems. We propose breakthrough acceleration techniques for a wide range of low-rank factorization-based first-order methods using GPUs, making the computation much more efficient and scalable. To illustrate the idea and effectiveness of our approach, we use the low-rank factorization-based SDP solver, LoRADS, as an example, which involves both the classic Burer-Monterio method and a novel splitting scheme with a starting logarithmic rank. Our numerical results demonstrate that the accelerated GPU version of LoRADS, cuLoRADS, can solve huge-scale semidefinite programming problems with remarkable efficiency. By effectively leveraging GPU computational power, cuLoRADS exhibits outstanding performance. Specifically, it can solve a set of MaxCut problems with 10^7 ×10^7 matrix variables in 10 seconds to 1 minute each on an NVIDIA H100 GPU with 80GB memory, whereas previous solvers demonstrated the capability of handling problems of this scale, required at least dozens of hours per problem on CPUs. Additionally, cuLoRADS shows exceptional scalability by solving 1) a MaxCut problem with a 170 million × 170 million matrix variable and 2) a Matrix Completion problem with a 20 million × 20 million matrix variable and approximately 200 million constraints, both in a matter of minutes.

2 - Simulation-Based Optimistic Policy Iteration In Identical-Interest Markov Games with KL Control Cost

Khaled Nakhleh, Texas A&M University, College Station, TX, United States, Ceyhun Eksin, Sabit Ekin

This work proposes an agent-based optimistic policy iteration (OPI) scheme for learning stationary optimal stochastic policies in identical interest Markov games, in which agents incur a Kullback-Leibler (KL) divergence cost for their control efforts.
This extends the KL optimal control framework for Markov decision process (MDPs) to multi-agent Markov games, which we call KL control identical-interest multi-agent games.
We introduce the Kullback-Leibler control optimistic policy iteration KLC-OPI scheme for learning an agent-based optimal policy.
KLC-OPI consists of a greedy policy improvement step, that is shown to follow a Boltzmann distribution, followed by an m-step temporal difference (TD) policy evaluation step.
KLC-OPI is shown to converge to the optimal stochastic policy asymptotically in the case of evaluating the entire state space (synchronous update), and when a subset of the state space is evaluated (asynchronous update) in each scheme iteration.

Based on the asynchronous m-step TD policy evaluation step of the scheme, we design an asynchronous, model-based, tabular multi-agent reinforcement learning algorithm, OPTIMAL: OPTimistic Policy Iteration for Multi-Agent Learning.
OPTIMAL evaluates the current policy using noisy m-step TD trajectories and updates the agents' value functions based on the unbiased noisy return.
Simulation results on a KL control variant of the Stag-Hare game validates our learning algorithm's performance in terms of minimizing the cost return.

3 - A Stability Principle for Learning under Non-Stationarity

Chengpiao Huang, Columbia University, New York, NY, United States, Kaizheng Wang

We develop a versatile framework for statistical learning in non-stationary environments. In each time period, our approach applies a stability principle to select a look-back window that maximizes the utilization of historical data while keeping the cumulative bias within an acceptable range relative to the stochastic error. Our theory showcases the adaptability of this approach to unknown non-stationarity. The regret bound is minimax optimal up to logarithmic factors when the population losses are strongly convex, or Lipschitz only. At the heart of our analysis lie two novel components: a measure of similarity between functions and a segmentation technique for dividing the non-stationary data sequence into quasi-stationary pieces.

4 - Sublinear Regret Bounds for An Actor-Critic Algorithm in Continuous-Time Linear-Quadratic Reinforcement Learning

Yilie Huang, Columbia University, New York, NY, United States, Yanwei Jia, Xunyu Zhou

We study reinforcement learning (RL) for continuous-time linear-quadratic (LQ) control where the volatility of the state process is both state- and control-dependent. We develop a model-free approach that relies neither on the knowledge of model parameters nor on their estimations, and devise an actor-critic algorithm to learn the optimal policy parameter. Our main contributions include the introduction of a novel exploration scheduling scheme, as well as a regret analysis of the proposed algorithm. We provide the convergence rate of the policy parameter to the optimal one, and prove that the algorithm achieves a regret bound of $O(N^{\frac{3}{4}})$. We conduct a simulation study to validate the theoretical results and demonstrate the effectiveness and reliability of the proposed algorithm. We additionally perform numerical comparisons of our algorithm against recent studies on model-based stochastic LQ RL algorithms adapted for state- and control-dependent volatility settings, demonstrating superior performance of our approach in terms of regret bounds.

TA11

Summit - 331

Data-driven Decisions: Insights and Impact Across Industries

Contributed Session

Chair: Yuan Qu, University of Hong Kong, Hong Kong, N/A, Hong Kong

1 - Analysis of the Impact of Fabricated Consumer Data on Corporate Profitability

bingchan Yang, Beijing Institute of Technology, Beijing, China, People's Republic of, Meichen Liu

Analysis of the Impact of Fabricated Consumer Data on Corporate Profitability

Abstact

This study aims to delve deeply into the multidimensional impact mechanism of consumer information falsification on corporate revenue. In the context of the information society, consumer data, as a critical resource, plays a vital role in determining the efficiency of information utilization and corporate revenue. However, consumers often provide false information due to privacy concerns, potentially leading to economic losses for enterprises. By constructing a theoretical model, this study systematically explores the relationship between privacy protection investment and the proportion of falsified consumer information, and further reveals how this relationship affects corporate revenue. The findings indicate that effective privacy protection investment can significantly reduce the proportion of false information, thereby enhancing the economic benefits of enterprises. Additionally, this study analyzes the specific impacts of various parameters, such as platform basic utility and the loss coefficient of information utilization, on corporate revenue, providing profound academic insights and practical guidance for enterprises in developing privacy protection and information management strategies.

KeyWord: Privacy Safeguarding /Corporate Earnings/ Information Misrepresentation

Bingchan Yang 1402588438@qq.com

2 - Redesigning Medicare Shared Savings Program: Implications for Risk-Averse ACOs

Dawei Jian, University of Wisconsin - Milwaukee, Milwaukee, WI, United States

The Medicare Shared Savings Program (MSSP) is designed to curb rising Medicare costs by incentivizing Accountable Care Organizations (ACOs) to provide healthcare more efficiently. However, faced with the obligation to assume financial risk, nearly two-thirds of ACOs withdrew from the MSSP. How should Medicare redesign MSSP to incentivize spending saving and encourage participation? We formulate the long-term MSSP through a dynamic principal-agent model, in which the Medicare provides a performance-based subsidy to partially reimburse ACOs' investments, considering their risk-aversion and the stochastic evolution of their private spending-saving abilities. Unlike the current cap-based system, we recommend lowering benchmarks and sharing rates initially, adjusting them over time based on the persistence of ACOs' abilities. Utilizing a dataset from CMS, we analyzed ACOs' risk attitudes to demonstrate how our approach could outperform the existing one.

3 - *Role of Individual differences in cybersecurity decision making in a workplace*

Sahar Farshadkhah, University of Illinois Springfield, Springfield, IL, United States

For several years, Ponemon Institute Data Breach Report showed that insiders are the weakest link in the cybersecurity management chain. Organizations increase their investment in cybersecurity drastically but still cybersecurity attacks costs significant amounts every year. It has never been more vital to do research on human-centric interventions. This study aims to explore the role of individual differences in explaining cybersecurity behaviors in the workplace. One of the main individual differences among employees could be their personality and how that affects their decision making while facing dilemmas related to complying with cybersecurity policies in a workplace. This study will look at this concept from a new angle and look for how individual differences and especially personality may change the known relationships among cybersecurity literature constructs. The findings has theoretical implications by extending the current literature and have practical implications for managers by helping them to have a better understanding of the role of individual differences in security-related behaviors and decision-making.

4 - Styling Box Selling to Strategic Buyers

Yuan Qu, University of Hong Kong, Hong Kong, Hong Kong, Max Z.J. Shen

Styling box selling is a new online shopping format rising recently. Instead of letting customers actively poodle around among numerous products on the platform, the styling box platform runs a passive shopping mode in which a customer will periodically receive a box of selected items. The customer can keep what they want and pay for them, then return everything else, where a discount will be applied if all the items are kept.

To explain the mechanism behind styling box selling, we take a holistic approach encompassing both a seller intent on maximizing profit and strategic buyers with heterogeneous preferences. First, we found that the rank of recommended items in the box, recommendation ordering, will significantly affect the profit structure. Under the homogeneous product setting, dominant recommendation ordering and discount strategies can be derived. Some monotonic dominance can also be ensured under the heterogeneous assumption. Then, we discuss a more general and realistic model with heterogeneous products and limited inventory, where the problem can be formulated as a Stackelberg knapsack game, an NP-hard bilevel integer programming. Besides recommendation ordering, we find that the structure of price ordering, the rank of price within all products, and valuation ordering, the rank of valuation to a product within all customers, provide unique properties in solving the problem.

TA12

Summit - 332

Human-Centric Decision-Making: From Psychosis to Sports Data

Contributed Session

Chair: Jinwook Lee, Drexel University

1 - Brain Speaks Louder Than Words

Qi Zhao, Penn State University, University Park, PA, United States, Elena Karahanna, D.J. Wu, Min Ding

Surveying participants is common for researchers and practitioners alike to capture perceptions, attitudes, and opinions. The validity of collected responses matters as they shape further managerial insights and theoretical findings. To avoid biases inherent in self-reports, this research leverages neuroscientific measurement and devises a novel approach to infer participants' true opinions directly from cognitive processing. The proposed approach draws upon cognitive dissonance theory that suggests that individuals engage in systematic cognitive processing when presented with an opinion that is incongruent with their own views. Leveraging this, we sequentially present artificial stimuli to participants and observe their cognitive processing through neuroscientific measurement. More systematic cognitive processing should occur when the artificial stimulus is incongruent with the participant's true opinion. For validation, we conduct a lab study to collect and infer participants' food preferences. We find significant neural evidence revealing more systematic processing upon incongruent stimuli exposure. We also examine how object familiarity and presence of a heuristic cue moderate the accuracy of our method. Additionally, we provide robustness evidence demonstrating the persistency of the induced neural response through multiple exposure repetitions. Contrary to existing methods, the proposed approach entails no training data at the individual level to make inferences and no direct reporting from participants, making itself also appealing to studies requiring a continuous participation experience.

2 - Smart Decision for a Better World

Anwer Ayyub, GTS, Dubai, United Arab Emirates

The decision-making process involves understanding the situation's reality, engaging stakeholders, and coordinating with various parties to identify potential issues, ensure the right project is chosen, and effectively communicated.

Appreciation and ideation are interconnected in decision-making, requiring diverse perspectives to effectively address internal and external stakeholders.

The data emphasizes the significance of considering external and internal evidence based, such as staff disengagement and a performance target culture and KPI, in business functions. It highlights the need for deliberate contemplation, understanding personality issues, and conflict resolution, as well as a big picture perspective for effective communication.

The phrase "co-creating" or "co-appreciating" emphasizes the importance of understanding a situation by putting oneself in the picture and allowing others to react, fostering collective understanding.

The PRISM framework emphasizes the importance of understanding a situation and problem, distinguishing between them, and breaking down actions into short-term and long-term ones. It helps businesses understand their profit-driven culture and focus on value creation. A successful business transformation project, canceled due to conflict, was achieved through a adoption of this framework.

References:

[1] Marquardt, M, J., 2014, Chapter 5: The Art of Asking Questions. In: Leading With Questions. San Francisco: Jossey-Bass, pp. 103-128.

[2] Winter, M., 2024, Study Guide 4: Practical Helps For Deciding EDMP Elective, Global MBA, AMBS, University of Manchester.

[3] Peter Check land, Sue Holwell. Information, systems and information systems: Making Sense of the field, International Journal of Information Technology Project Management, 13:1, (1-20), Online publication date: 01-Feb-1998

3 - A Reinforcement Learning-based Automated Negotiation Model for Sellers in the Supply Chain

Ga Hyun Lee, Kyung Hee University, Yongin-si, Korea, Republic of, Hyun Woo Jeon

The negotiation is a dynamic decision-making process with a lot of complex information to consider. Human negotiators, however, can have difficulties in making the optimal decision due to various limitations related to time, cost, and information processing capabilities. In addition, in the process of negotiation in the supply chain, sellers have to deal with bargaining power asymmetry, because they are typically small and medium-sized enterprises (SMEs). To address these issues, this study proposes an automated negotiation model from the perspective of sellers. In particular, we develop a reinforcement learning (RL) based model utilizing the Q-learning algorithm as RL technology is widely utilized to develop automated negotiation models. In addition, to deal with the insufficient amount of data, we simulate the multiple negotiation episodes based on the correlation structures of real transaction data; quantity (Q), price (P), and delivery lead time (D). Specifically, we generate more than 20,000 episodes for two negotiation cases: (i) negotiating P dependent on Q and (ii) negotiating P dependent on D. Then we suggest the model training results of each negotiation case. The results show that the developed model in this study can be applied to various negotiation situations and helpful for sellers in the supply chain.

4 - Decoding Decision-Making Processes: Bottom-Up Tree Construction through Model-Based Methods with Consideration of Data-Driven Approaches

Jinwook Lee, Drexel University, Philadelphia, PA, United States, LANQING DU, Sung-Cheol Kim, Matthew Schneider

Human decision-making inherently involves subjectivity. People often exhibit tendencies to favor those within their social circles, leading to differing treatment of individuals. Determining privilege, representation, and other relevant factors is a complex task. This research seeks to develop models for conducting comparative analyses with reduced subjectivity. We employ two main methodologies: model-based and data-driven approaches. In the model-based approach, we utilize set-theoretic operations to examine differences and similarities among groups, ensuring comprehensive coverage through subgroup division. Meanwhile, the data-driven approach involves calculating the importance of features using SHAP values to identify significant variables for each group. To validate these methods, we conduct extensive numerical experiments using a recent dataset comprising 51 judicial circuits and over 100,000 criminal justice outcomes from the Department of Community Supervision (DCS) in the State of Georgia.

TA13

Summit - 333

Supply Chain Risk and Resilience

Contributed Session

Chair: Qiang Qiang, Penn State, Malvern, PA, United States

1 - Synergy or competition: how does the working capital of supply chain network members affect the performance of enterprises?

Mengyin Li, China Conservatory of Music, Beijing, China, People's Republic of, Hua Song, Xinge Ding

The operation of enterprise is influenced by its supply chain network. How does supply chain network working capital management affect firm performance? On the one hand, companies can benefit from shifting inventory and accounts pressures to upstream and downstream; on the other hand, improving the efficiency of the entire supply chain network can also bring benefits to enterprises. This study verifies the impact of inter-organizational capital efficiency on firm performance. This paper takes 214 listed manufacturing companies that disclose upstream and downstream information of their supply chains as research samples, and collects 1070 observations, as well as information from 13638 suppliers and 8562 customers, to conduct empirical research on supply chain network and working capital management. Research shows that companies benefit more when the average cash conversion cycle of all participating partners across the supply chain network is reduced.The synergy effect among supply chain networks is more significant than the competition effect.

2 - Dual-sourcing of capacity: Frequent batching or more on-demand?

Partha Mishra, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Sunil Chopra, Sebastien Martin, Karen Smilowitz

Dual-sourcing, a strategy that diversifies supply sources to manage uncertainty, is widely used in supply chain management. We extend this concept to the contexts of transportation and manufacturing, introducing the notion of "dual-sourcing of capacity." In these contexts, we model the trade-off between cost and responsiveness using two capacity sources: a cheaper batched source operating at a fixed frequency (e.g., shuttles or furnaces) and a more expensive on-demand source with unit capacity (e.g., Ubers or heat lamps). Unlike traditional dual-sourcing, which focuses on the supply of goods, we use the term sourcing to mean using a resource's capacity to perform a task, e.g., using vehicles to transport passengers or using heating equipment to process semiconductor wafers. We prove that a simple threshold policy, which we call "Q-policy", is optimal under certain restrictions. Lowering the threshold Q in Q-policies increases on-demand use in our setting; however, we find it is never optimal to reduce Q below the batched source's batch size. Furthermore, when the optimal Q is unchanged, greater responsiveness is achieved by increasing the frequency of using the batched source, which reduces the on-demand usage, challenging conventional wisdom. Our findings from the fresh perspective of dual-sourcing provide useful insights for managing uncertainty in transportation and manufacturing contexts.

3 - The Multi-period Interactive CLSC Network

Qiang Qiang, Penn State, Malvern, PA, United States

Inspired by the recent business cases, we develop a coupled multi-period CLSC network model dealing with heterogeneous products facing different demand markets. The end-of-life product from the forward supply chain is collected, recycled and the raw material is extracted to be used as an input for the reverse supply chain which produces another type of product. In particular, interactions between the two supply chains with inventory cost and innovation competition are analyzed. An algorithm is introduced to study the network equilibrium. Numerical examples are used to illustrate the model.

TA14

Summit - 334

Financial Engineering and Technology

Invited Session

Finance

Chair: Xuedong He, The Chinese University of Hong Kong, Shatin, Hong Kong

1 - A tail-risk sensitive reinforcement learning approach for option hedging

Xianhua Peng, Peking University, Shenzhen, China, People's Republic of, Xiang Zhou, Bo Xiao, Yi Wu

We propose a new risk-sensitive reinforcement learning approach for the dynamic hedging of options and other derivatives. The approach focuses on the minimization of the tail risk of the final P&L of the seller of options. Different from most existing reinforcement learning approaches that require a parametric model of the underlying asset, our approach can learn the optimal hedging strategy directly from the historical market data without specifying a parametric model. Our approach presents an efficient way to incorporate the tail risk measures of the final P&L into traditional reinforcement learning framework in a model-free setting. We carry out comprehensive empirical study to show that, in the out-of-sample tests, the proposed reinforcement learning hedging strategy can obtain statistically significantly lower tail risk and higher mean of the final P&L than traditional delta hedging methods.

2 - Systemic Values-at-risk: Computation and Convergence

Cagin Ararat, Bilkent University, Ankara, Turkey, Mustafa Çolak

We investigate the convergence properties of sample-average approximations (SAA) for set-valued systemic risk measures. We assume that the systemic risk measure is defined using a general aggregation function with some continuity properties and value-at-risk applied as a monetary risk measure. Our focus is on the theoretical convergence of its SAA under Wijsman and Hausdorff topologies for closed sets. After building the general theory, we provide an in-depth study of an important special case where the aggregation function is defined based on the Eisenberg-Noe network model. In this case, we provide mixed-integer programming formulations for calculating the SAA sets via their weighted-sum and norm-minimizing scalarizations. To demonstrate the applicability of our findings, we conduct a comprehensive sensitivity analysis by generating a financial network based on the preferential attachment model and modeling the economic disruptions via a Pareto distribution.

3 - Index Investing and Dynamic Information Environment

Yiwen Shen, HKUST, Hong Kong, Hong Kong, Yueting Jiang

In this paper, we study how index investing and dynamic information environment affect market volatility, correlation, and price informativeness. We consider an overlapping generation model with multiple risky assets.The market has two classes of investors, with the first type trading single stocks and the second type trading an index fund. The single-stock investors have access to the private signal of the stock they trade, while the index investors rely on public information. The information environment varies in response to the investors' composition, capturing a feedback effect that more single-stock investors increase the amount of available information. We solve the investors' participation choice and optimal demands that clear the market under a framework of rational expectation equilibrium. Our model reveals the impact of indexing, information asymmetry, and stock correlation on the market equilibrium.

4 - Data-Driven Option Pricing

Xi Yang, National University of Singapore, Singapore, Singapore, Min Dai, Hanqing Jin

We propose an innovative data-driven option pricing methodology that relies exclusively on the dataset of historical underlying asset prices. While the dataset is rooted in the objective world, option prices are commonly expressed as discounted expectations of their terminal payoffs in a risk-neutral world. Bridging this gap motivates us to identify a pricing kernel process, transforming option pricing into estimating expectations in the objective world. We recover the pricing kernel by solving a utility maximization problem, and estimate the expectations in terms of a functional optimization problem. Leveraging the deep learning technique, we design data-driven algorithms to solve both optimization problems over the dataset. Numerical experiments are presented to demonstrate the efficiency of our methodology.

5 - Portfolio Selection with Time-Varying Taxation

Xianhao ZHU, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Chen Yang

We present an optimal control framework for the continuous-time portfolio selection problem under time-varying capital gains taxes. Our study focuses on the investor’s decision in anticipation of tax rate adjustments. Our results suggest that, in contrast to a constant tax rate, investors tend to realize capital gains while deferring a significant proportion of capital losses in response to a tax rise. The magnitude of the tax rate increment determines the extent of capital gains realization. Conversely, investors refrain from realizing capital gains as the date of a tax cut approaches. We also examine scenarios where the direction of the tax rate is uncertain, as well as cases involving an intermediate period before the adjusted tax rate takes effect. Our results align with the empirical findings and contribute to the understanding of market behavior preceding a capital tax rate adjustment.

TA15

Summit - 335

Innovative Applications in Revenue Management & Pricing

Invited Session

Revenue Management and Pricing

Chair: Nur Sunar, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States

Co-Chair: Rahul Roy, Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Co-Chair: Jayashankar Swaminathan, University of North Carolina Chapel Hill, Chapel Hill, NC, United States

1 - Asymptotic Optimality of Dual Base-Stock Policies for Omni-Channel Perishable Inventory Systems

Yin Pan, Zhejiang University, Hangzhou, China, People's Republic of, Sentao Miao, Jinzhi Bu, Mingzheng Wang

We consider periodic review omni-channel inventory systems where there are an online channel and an offline channel selling the same perishable product with a fixed product lifetime over an infinite planning horizon. In online channel, it sells the product with a shorter lifetime and the outdated product goes to the offline channel. Unsatisfied demand can be either lost or backlogged. The objective is to minimize the long-run average holding, penalty, and outdating cost. The optimal policy for perishable system is notoriously complex and computationally intractable because of the curse of dimensionality. While there exist asymptotic results in the literature for single-channel perishable inventory systems, no paper has specifically focused on omni-channel perishable inventory systems. In this paper, we first focus on the omni-channel perishable system under a first-in-first-out issuance policy and show that a simple dual base-stock policy is asymptotically optimal when any one of the product lifetime, unit penalty cost, and unit outdating cost becomes large; moreover, its optimality gap converges to zero exponentially fast in the first parameter. We then derive the same theoretical results for the omni-channel perishable system under a last-in-first-out issuance policy. Finally, we provide a numerical study to demonstrate the performances of dual base-stock policies in these systems.

2 - Learning to Price Supply Chain Contracts against a Learning Retailer

, Xuejun Zhao, Ruihao Zhu, William Haskell

In this paper, we study the supply chain contract design problem faced by a data-driven supplier (she) who needs to respond to the inventory decisions of the downstream retailer (he). Both the supplier and the retailer are uncertain about the market demand and need to learn about it sequentially over a fixed time horizon. In addition, the supplier does not know the retailer's inventory learning policy, which may change dynamically. The goal for the supplier is to develop data-driven pricing policies with sublinear regret bounds under a wide range of possible retailer inventory learning policies. We establish a connection with nonstationary online learning by following the notion of a variation budget. We start by making the observation that existing approaches for non-stationary online learning cannot precisely delineate the dynamics incurred by the retailer's inventory learning policy. To overcome this challenge, we introduce a new notion of variation budget, which better quantifies the impact of the retailer's learning on the supplier's decision-making environment. We then proceed to propose dynamic pricing policies for the supplier for both discrete and continuous demand distributions, which lead to sublinear regret bounds for the supplier. Our pricing policies empirically outperform those from the existing non-stationary learning literature. At the managerial level, we answer affirmatively that there is a pricing policy with a sublinear regret bound for the supplier under a wide range of retailer inventory learning policies, even though she faces a learning retailer and an unknown demand distribution.

3 - Robust Optimal Selling Mechanism under Non-linear Utility

Yicheng Liu, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Pin Gao, Shixin Wang, zhen wang, Zizhuo Wang

We consider a robust mechanism design problem in which a seller wishes to choose a selling mechanism to maximize the revenue obtained from a buyer with private willingness-to-pay information and a nonlinear utility function. The objective is to provide a profit guarantee for the seller across all possible willingness-to-pay distributions whose mean and variance are consistent with the prior information. When the utility function is of a quadratic form and the coefficient of variation is low, we find that a quadratic pricing mechanism yields the optimal revenue guarantee. In this case, the worst-case distribution of willingness-to-pay follows a Pareto distribution. However, when the coefficient of variation is high, a unit price mechanism is optimal. We further extend our analysis to general elasticity utility functions and investigate how the optimal worst-case revenue varies with elasticity and coefficient of variation. Furthermore, given the widespread implementation of a two-part tariff mechanism in real-world scenarios, we also study the optimal two-part tariff mechanism. We show that under the quadratic utility form, the optimal two-part tariff pricing strategy achieves at least 80% of the optimal worst-case revenue when compared with the optimal robust mechanism. We also develop corresponding performance guarantees when the elasticity coefficient is less than two. However, when the elasticity coefficient exceeds two, the competitive ratio of the two-part pricing scheme goes to zero as the variance of willingness-to-pay grows to infinity.

4 - Peer Review Market Design

Craig Fernandes, University of Toronto, Toronto, ON, Canada, James Siderius, Raghav Singal

Peer review is a cornerstone of academic research, but its effectiveness depends on reviewers exerting costly effort to evaluate papers. This study addresses how peer review markets can be designed to incentivize high levels of reviewing effort and better identify paper quality. We analyze a one-sided market where each agent acts as both an author and a reviewer, with the central designer (e.g., conference organizer) aiming to maximize the number of high-quality submissions accepted by employing various matching mechanisms. We first show that the random matching mechanism (similar to current practice) can be arbitrarily bad as it always induces zero reviewing effort. Second, we study a recently proposed symmetric matching mechanism that pairs high-quality reviewers with other high-quality reviewers (and similarly for low-quality reviewers). While this approach is an improvement over random, it can be exploited by low-quality authors who prefer lower quality (and thus noisier) reviews, leading to bad market outcomes. To address these challenges, we propose a new admission control mechanism, which rejects papers from reviewers who fail to exert effort. Though stringent, this mechanism achieves the first-best welfare. We explore several extensions where effort levels are noisily observed and where authors upload preprints to disclose their reputation, finding consistent results across these variations. We conclude with practical insights for enhancing the effectiveness of peer review. To the best of our knowledge, we are the first to understand the value of operational levers such as matching and admission control in the design of peer review markets.

TA16

Summit - 336

Dynamic Assortment and Pricing Optimization

Invited Session

Revenue Management and Pricing

Chair: Xu Sun, University of Miami, Miami, FL, United States

Co-Chair: Cong Shi, University of Miami, 5250 University Dr, Miami, FL, 33146-2000, United States

1 - Demand Balancing in Primal-Dual Optimization for Blind Network Revenue Management

Yining Wang, University of Texas at Dallas, Richardson, TX, United States, Sentao Miao

This paper proposes a practically efficient algorithm with optimal theoretical regret which solves the classical network revenue management (NRM) problem with unknown, nonparametric demand. Over a time horizon of length T, in each time period the retailer needs to decide prices of N types of products which are produced based on M types of resources with unreplenishable initial inventory. When demand is nonparametric with some mild assumptions, Miao and Wang (2021) is the first paper which proposes an algorithm with O(poly(N,M,ln(T))T‾‾√) type of regret (in particular, Õ (N3.5T‾‾√) plus additional high-order terms that are o(T‾‾√) with sufficiently large T≫N). In this paper, we improve the previous result by proposing a primal-dual optimization algorithm which is not only more practical, but also with an improved regret of Õ (N3.25T‾‾√) free from additional high-order terms. A key technical contribution of the proposed algorithm is the so-called demand balancing, which pairs the primal solution (i.e., the price) in each time period with another price to offset the violation of complementary slackness on resource inventory constraints. Numerical experiments compared with several benchmark algorithms further illustrate the effectiveness of our algorithm.

2 - Regularized Predict-Then-Optimize for Pricing with Unknown Demand Functions

Guan Wang, Rotman School of Management, Toronto, ON, Canada, Michael Albert, Max Biggs, Ningyuan Chen

<div full-size"=""><div pdfjs-viewer-outer"="">

In data-driven pricing, the firm is presented with a data set with price/demand pairs and an unknown linear demand function. The firm needs to design a price based on the data set to maximize the revenue.
We develop a novel framework to regularize the “predict-then-optimize” (PTO) approach. Our method is remarkably straightforward: we first obtain the Ordinary Least Squares (OLS) estimator, and then adjust it by multiplying with a data-driven factor, λ, where λ > 1. As a result, the proposed price is less than the price obtained from PTO. We show that this price always outperforms the price obtained from PTO in terms of the expected revenue when the data size increases. Our numerical results further confirm that this regularized price surpasses the PTO price even in scenarios with limited data. Additionally, unlike traditional
regularization approaches in linear models such as ridge regression or LASSO, our regularization inflates β instead of shrinking it toward zero. It suggests the complexities introduced by the optimization phase in the classical estimation process.

3 - Advanced Reservation Problem with Reusable Resource Under Two TIME Scales

Yun-Tung Kuo, Duke University, Durham, NC, United States, Yehua Wei, Cong Shi

We study a general online reservation model, where a decision maker attempts to maximize the total revenue endowed with limited reusable resources by accepting or rejecting advanced reservation requests. In our model, customer requests arrive continuously over time, but services to start and end on a daily basis. For this two time scales model, we design an online algorithm for the decision maker and show that the algorithm achieves asymptotic optimality under mild assumptions.

4 - Beyond Profit: Resource Pricing for Social Impact and Financial Viability

Xu Sun, University of Miami, Miami, FL, United States

Motivated by social enterprises (SE) that offer affordable and scalable reusable products or services, we consider a dynamic pricing problem faced by an SE managing various resources, each with a fixed reusable capacity or initial inventory. Price-sensitive customers arrive over time, each requesting a unit of a specific resource. When the rental ends, the resource is released and becomes available for future demand. The SE aims to optimize long-term social value while ensuring that the long-run average revenue meets or exceeds a specified threshold. We develop heuristics that solve the problem asymptotically. Numerical experiments show that the proposed solutions are highly effective.

TA17

Summit - 337

Learning and Incentives

Invited Session

Revenue Management and Pricing

Chair: Meng Qi, Cornell University, Ithaca, NY, United States

Co-Chair: Mingxi Zhu, Georgia Institute of Technology, Atlanta, GA, 30309, United States

1 - Mechanism for Decision-Aware Collaborative Federated Learning: a Pitfall of Shapley Values

Meng Qi, Cornell University, Ithaca, NY, United States, Mingxi Zhu

This paper investigates mechanism design for decision-aware collaboration via federated learning (FL) platforms. Our framework consists of a digital platform and multiple decision-aware agents, each endowed with proprietary data sets. The platform offers an infrastructure that enables access to the data, creates incentives for collaborative learning aimed at operational decision-making, and conducts FL to avoid direct raw data sharing. The computation and communication efficiency of the FL process is inherently influenced by the agent participation equilibrium induced by the mechanism. Therefore, assessing the system's efficiency involves two critical factors: the surplus created by coalition formation and the communication costs incurred across the coalition during FL. To evaluate the system efficiency under the intricate interplay between mechanism design, agent participation, operational decision-making, and the performance of FL algorithms, we introduce a multi-action collaborative federated learning (MCFL) framework for decision-aware agents. Under this framework, we further analyze the equilibrium for the renowned Shapley value based mechanisms. Specifically, we examine the issue of false-name manipulation, a form of dishonest behavior where participating agents create duplicate fake identities to split their original data among these identities. By solving the agent participation equilibrium, we demonstrate that while Shapley value effectively maximizes coalition-generated surplus by encouraging full participation, it inadvertently promotes false-name manipulation. This further significantly increases the communication costs when the platform conducts FL. Thus, we highlight a significant pitfall of Shapley value based mechanisms, which implicitly incentivizes data splitting and identity duplication, ultimately impairing the overall efficiency in FL systems.

2 - Platforms for Efficient and Incentive-Aware Collaboration

Mingda Qiao, University of California, Berkeley, Berkeley, CA, United States, Nika Haghtalab, Kunhe Yang

Collaboration is crucial for reaching collective goals. However, its effectiveness is often undermined by the strategic behavior of individual agents -- a fact that is captured by a high Price of Stability (PoS) in recent literature [Blum et al., 2021]. Implicit in the traditional PoS analysis is the assumption that agents have full knowledge of how their tasks relate to one another. We offer a new perspective on bringing about efficient collaboration among strategic agents using information design. Inspired by the growing importance of collaboration in machine learning (such as platforms for collaborative federated learning and data cooperatives), we propose a framework where the platform has more information about how the agents' tasks relate to each other than the agents themselves. We characterize how and to what degree such platforms can leverage their information advantage to steer strategic agents toward efficient collaboration.

Concretely, we consider collaboration networks where each node is a task type held by one agent, and each task benefits from contributions made in their inclusive neighborhood of tasks. This network structure is known to the agents and the platform, but only the platform knows each agent's real location. We employ private Bayesian persuasion and design two families of persuasive signaling schemes that the platform can use to ensure a small total workload when agents follow the signal. The first family aims to achieve the minmax optimal approximation ratio compared to the optimal collaboration. The second family ensures per-instance strict improvement compared to full information disclosure.

3 - Estimating and Incentivizing Imperfect-Knowledge Agents with Hidden Rewards

Ilgin Dogan, INSEAD, Fontainebleau, France, Zuo-Jun Shen, Anil Aswani

As opposed to many principal-agent models, incentive providers in practice often cannot observe the reward realizations of incentivized agents. This information asymmetry challenges the principal to lead the selfish agent by solely watching their actions, which becomes even more challenging when these actions subject to the agent's learning uncertainty. We explore this practically relevant scenario of repeated adverse selection game within a multi-armed bandit framework and jointly address: i) the consistent estimation of the agent's rewards from bounded continuous spaces, and ii) the provision of adaptive, data-driven incentives that achieve low-regret for the principal.

4 - Building Collaborative Intelligence

Sai Praneeth Reddy Karimireddy, UC Berkeley, Berkeley, CA, United States

How do we build ML systems that put the interests of users and society front and center? This talk will explore the potential of **collaborative learning (CL)** for such user-centric ML, drawing on some pilot projects with Doctors Without Borders and the Cancer Registry of Norway. We show how CL techniques such as decentralized and federated learning can allow us to train ML models without the users sacrificing their privacy or ownership and control over their data. Yet for these systems to truly succeed, two fundamental challenges must be confronted. These systems need to be efficient and **scale to massive networks,** and manage and resolve the **conflicting goals** of the participants. We will discuss how tools from optimization, statistics, and economics can be leveraged to address these challenges.

TA18

Summit - 338

Optimization Under Uncertainty: Pricing, Inventory, and Learning

Invited Session

Revenue Management and Pricing

Chair: Jinzhi Bu, The Hong Kong Polytechnic University, Hong Kong, N/A

Co-Chair: Lei Li, Hong Kong Polytechnic University, M507n, The Hong Kong Polytechnic University, Hong Kong, N/A

1 - Randomized Robust Price Optimization

Xinyi Guan, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Velibor Misic

The robust multi-product pricing problem is to determine the prices of a collection of products so as to maximize the worst-case revenue, where the worst case is taken over an uncertainty set of demand models that the firm expects could be realized in practice. A tacit assumption in this approach is that the pricing decision is a deterministic decision: the prices of the products are fixed and do not vary. In this paper, we consider a randomized approach to robust pricing, where a decision maker specifies a distribution over potential price vectors so as to maximize its worst-case revenue over an uncertainty set of demand models. We formally define this problem – the randomized robust price optimization problem – and analyze when a randomized price scheme performs as well as a deterministic price vector, and identify cases in which it can yield a benefit. We also propose two solution methods for obtaining an optimal randomization scheme over a discrete set of candidate price vectors based on constraint generation and double column generation, respectively, and show how these methods are applicable for common demand models, such as the linear, semi-log and log-log demand models. We numerically compare the randomized approach against the deterministic approach on a variety of synthetic and real problem instances; on real data instances derived from a grocery retail scanner dataset, the improvement can be as high as 92%.

2 - Optimizing Rarity: the Economics of Blind Boxes with Random Items

TAOJIE QIN, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Shreyas Sekar

Blind box selling is a novel selling mechanism where buyers purchase sealed packages containing unknown items, with the chance of uncovering rare or special items. To unravel the economic motivation behind blind boxes, we propose a stylized model encapsulating the core trade-offs sellers face when designing a blind box package, including price and drawing probability. The model distinguishes between two types of items: regular and special, with the latter valued higher by customers due to its rarity. The customers' marginal utility from opening the blind box decreases with each unique item received, whether regular or special, but remains unchanged if they receive a duplicate item. Naturally, customers cease purchasing when the expected marginal utility becomes negative. We compare blind box selling with traditional separate selling in terms of optimal profit, sales volume, and price. Our findings indicate that item heterogeneity and rarity are necessary for the success of blind box selling. Surprisingly, we show that blind boxes are more profitable in extreme scenarios where the manufacturing costs are either low or high. We also find that blind boxes perform better when customers are highly status-seeking. However, our study also illustrates that blind boxes are not a panacea and that the seller has to deploy them carefully, especially when manufacturing costs are neither high nor low. To examine the robustness of our results, we consider cases such as endogenous supply and customer heterogeneity.

3 - Learning in Lost-Sales Inventory Systems with Stochastic Lead TIMES and Random Supplies

Jiameng Lyu, Tsinghua University, Beijing, China, People's Republic of, Xin Chen, Shilin Yuan, Yuan Zhou

Supply uncertainty, characterized by stochastic lead times and random supply quantities, has attracted increasing attention from academia, industries, and governments, particularly in the aftermath of the COVID-19 pandemic. In this paper, we consider the problem of managing lost-sales inventory systems with general supply uncertainty: stochastic lead times and random supplies. Unlike the previous studies, we assume the decision maker has no prior information on the stochastic demand and supply. We propose the first provably effective learning algorithm for inventory management problems with censored demand and supply data under general supply uncertainty. Then we establish a regret of $O(L+\mathrm{poly}(L')\sqrt{T})$ for this learning algorithm compared to the best constant-order policy, where L’ is the upper bound of the random part and L is the deterministic part of the stochastic lead times. Due to the complicated nature of the considered inventory systems, this problem exhibits three primary technical challenges: the non-convexity of cost function, the establishment of stability for inventory systems under constant-order policies, and the accurate estimation of long-run average costs. We overcome these challenges through novel approaches, some of which are of independent interest. We also conduct numerical experiments to demonstrate the effectiveness of our algorithm.

4 - Bayesian Dynamic Pricing and Subscription Period Selection with Unknown Customer Utility

Yuan-Mao Kao, Baruch College, The City University of New York, New York City, NY, United States, Bora Keskin, Kevin Shang

We consider a provider offering a subscription service to customers over a planning horizon. The customers decide whether to subscribe according to a utility model. The provider has a prior belief about the customer utility model and updates its belief based on the transaction data of new customers and the usage data of existing subscribers. The provider aims to minimize its regret - the expected profit loss relative to a clairvoyant who knows the customer utility model. To analyze regret, we first study the clairvoyant's full-information problem, noting that the resulting dynamic program suffers from the curse of dimensionality. We characterize the optimal policy for the full-information problem via a customer-centric approach that balances the provider's immediate and future profits from a customer. When the provider does not have full information, we find that the simple and commonly used certainty-equivalence policy exhibits poor performance. We illustrate that this can be due to incomplete or slow learning but can also occur because of offering a suboptimal contract with a long subscription period in the beginning. We develop an adaptive learning policy, namely the information-threshold policy, that focuses on learning until the provider's accumulated information exceeds a chosen threshold. We show that this policy achieves asymptotically optimal performance with its regret growing logarithmically in the planning horizon. Our results indicate that offering a long subscription period could be costly when the provider knows little about the customers' usage and the service cost is uncertain.

5 - Online Learning Algorithms for Multi-product Inventory Systems with Shipping Costs

Liutao YANG, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Shining Wu

We consider the inventory optimization problem of an e-commerce firm that sells two products with unknown correlated demand where sequentially arriving customers may demand either or both products. The firm uses partial fulfillment in the way that items requested in an order are shipped via one package if they are all available or are delivered by multiple sequential shipments otherwise. We formulate the problem, analyze the optimal inventory decisions when the firm has full demand information, and develop machine learning-based algorithms for online optimization of the inventory decisions when the information is unknown and needs to be learned. These algorithms are proven efficient and effective, with a tight regret bound of $\Omega(\sqrt{T})$. We also conduct numerical experiments, highlighting the impact of incorporating shipping costs into inventory decision-making.

TA19

Summit - 339

Solar PV Innovations and Impacts

Contributed Session

Chair: Yagmur Arioz, 182 Park Ave, 01609

1 - Bi-LSTM for PV Power Forecasting: A Novel Approach for Grid Integration with EV Charging

Mengyao He, IE Business School - IE University, Madrid, Spain, Konstantina Valogianni

Photovoltaic power is one of the widely deployed renewable energy resources. However, the climate-dependent nature of photovoltaic (PV) power makes the energy generation intermittent and volatile. This intermittency could be problematic, especially for those who need a stable and reliable grid. In addition, this renewable energy could be coupled with electric vehicle charging, leading to reduced CO_2 emissions. We proposes a novel approach for forecasting photovoltaic (PV) generation using Bi-directional Long Short-Term Memory (Bi-LSTM) neural networks. Bi-LSTM's strength in handling sequential data, such as time-series data and textual data, allows for accurate predictions of PV power based on historical weather conditions. Furthermore, we explore the integration of a stochastic optimization electric vehicle (EV) charging model with the prediction system. This combined framework aims to achieve a balance between maximizing renewable energy usage and offering flexibility in EV charging schedules.

2 - Advanced Signal Decomposition Analysis and Anomaly Detection in Photovoltaic Systems

mahya qorbani, Georgia Institute of Technology, Atlanta, GA, United States, Kamran Paynabar

With the rapid expansion of large-scale photovoltaic (PV) plants, it is paramount for solar stakeholders to understand the reliability and efficiency of their plants to inform maintenance decisions, increase production, and understand the design factors that impact performance. Diagnosing underperformance in PV plants is challenging due to the relatively few monitoring points with respect to the large geographic footprint of the plant. This study introduces a cutting-edge method that transforms the analysis and management of key factors influencing PV plant performance, including performance loss rate (PLR), recoverable soiling, and major system changes. Identifying these factors is critical for deriving actionable insights. Leveraging advanced analytical techniques such as wavelet transformation, robust regression, and extreme point analysis, this approach provides a nuanced understanding of these factors. This method has been tested across two synthetic datasets and one real dataset, consistently surpassing existing benchmarks by achieving a lower median mean absolute error and reduced error variability across all comparable components.

3 - Balancing Energy Access and Clean Energy Access: Evaluating The Marginal Benefits Of Rooftop Solar PV Policies

Afsal Najeeb, IIT Bombay, Mumbai, India, Anand Rao

Policy incentivesaimed at encouraging offtake of Rooftop Solar PV (RSPV) systems by residential consumers are common in most countries. These incentives operate in the midst of policies on pricing, tariff subsidies and varying levels of solar yield in different parts of the country. The paper attempts to understand the economic viability of Rooftop Solar PV (RSPV) systems for residential consumers in India. Put simply, it aims to answer if it makes economic sense for a residential consumer to install RSPV systems and understand the conditions under which the investment becomes profitable. Load profiles of consumers drawn from different parts of the country with different levels of self consumption were obtained using smart metering data. The analysis shows that state level policies like the price of electricity, feed in tariff and availability of subsidies play a critical role in determining the economic viability of RSPV systems. An interesting effect of conflicting effects of energy equity policies and renewable energy promotion policies are observed. We find that optimally sized RSPV systems can lead to significant net present economic benefits of up to USD 1800 per year and a reduction of cost of levelized per unit cost of electricity by up to 45%. Capital subsidy remains critical to the economic viability of RSPV systems, even at high levels of consumption. Lack of proportionate benefits of export of electricity indicate that the present policy environment encourages offsetting of self-consumption with self-generation but stops short of encouraging export of electricity for economic benefit.

4 - Poverty Alleviation through Solar Energy Systems: Bridging Gaps for a Sustainable Future

Yagmur Arioz, Ankara Yıldırım Beyazıt University, Ankara, Turkey, Abdullah YILDIZBASI, Joseph SARKIS

Solar energy systems improve the quality of life by providing access to clean energy, stimulating economic development, and alleviating poverty, thereby contributing to the achievement of sustainable development goals. This study deals with the co-benefit theory and ecological modernization theory, which offer different perspectives for understanding the impacts and strategies related to the integration of solar energy into efforts aimed at poverty alleviation. By providing a reliable and renewable energy source, solar energy systems can reduce energy costs for low-income households, improve health outcomes by decreasing reliance on harmful traditional fuels, and enhance educational opportunities through better lighting and access to technology. Furthermore, the adoption of solar energy systems can create job opportunities in installation, maintenance, and manufacturing, thereby fostering economic growth. Despite these benefits, there are several critical gaps that hinder the widespread adoption and effectiveness of solar energy systems in poverty alleviation. This study addresses these gaps in the implementation of solar energy systems for poverty alleviation.

5 - Circular Economy in Solar PV: Analyzing the 10Rs through Stakeholder Theory

Yagmur Arioz, Ankara Yıldırım Beyazıt University, Ankara, Turkey, Abdullah YILDIZBASI, Ibrahim Yilmaz

The solar PV supply chain has recently surged in importance for global clean energy transitions, crucial for meeting sustainable development goals. This shift is driven by the principles of the circular economy, which are integral for enhancing energy efficiency, prolonging the life of solar panels, reducing waste, minimizing value loss, and achieving sustainability targets. To meet these goals, it's vital to improve stakeholder interactions, communication, and collaboration within the solar PV supply chain. However, there is a noticeable gap in the literature regarding the evaluation of the 10R circularity targets within the solar PV supply chain from the perspective of stakeholders. This study aims to address this gap by examining the roles and interactions of solar PV stakeholders in achieving circularity goals, offering managerial insights through the lens of stakeholder theory and employing a hybrid multi-criteria decision-making approach based on a real-world case. The results indicate that stakeholder theory is a robust framework for supporting the development and sustainability of circular supply chains in the solar PV sector.

TA20

Summit - 340

Data-Driven Approach to Decision Support

Invited Session

Decision Analysis Society

Chair: Eunbi Kim, Korea University., Nowon-gu, Seoul

Co-Chair: Joonyup Eun, Korea Univeristy, Seoul, Korea, Republic of

1 - Real-time Adaptation for Time-series Signal Prediction using Label-aware Neural Processes

Seokhyun Chung, University of Virginia, Charlottesville, VA, United States, Raed Al Kontar

Building a predictive model that rapidly adapts to real-time condition monitoring (CM) time-series data is critical for engineering systems/units. Unfortunately, many current methods suffer from a trade-off between representation power and agility in online settings. For instance, parametric methods that assume an underlying functional form for CM signals facilitate efficient online prediction updates. However, this simplification leads to vulnerability to model specifications and an inability to capture complex signals. On the other hand, approaches based on over-parameterized or non-parametric models can excel at explaining complex nonlinear signals, but real-time updates for such models pose a challenging task. In this paper, we propose a neural process-based approach that addresses this trade-off. It encodes available observations within a CM signal into a representation space and then reconstructs the signal's history and evolution for prediction. Once trained, the model can encode an arbitrary number of observations without requiring retraining, enabling on-the-spot real-time predictions along with quantified uncertainty and can be readily updated as more online data is gathered. Furthermore, our model is designed to incorporate partial information on qualitative factors (e.g., missing labels) from individual units. This integration not only enhances individualized predictions for each unit but also enables joint inference for both signals and their associated labels. Numerical studies on both synthetic and real-world data in degradation modeling highlight the advantageous features of our model in real-time adaptation, enhanced signal prediction with uncertainty quantification, and joint prediction for labels and signals.

2 - A data-driven approach to improving student perceptions of undergraduate analytics courses

Baback Vaziri, James Madison University, Harrisonburg, VA, United States, Jacob Sanzano, Luis Novoa, Gizem Atav

As educators, we have different approaches to improve student experiences in our classroom. Oftentimes, however, experiences can be shaped by their perceptions going into the class. In this study, we collect data to understand perceptions of students coming into an undergraduate business course in analytics – specifically, management science. Given the rising importance of analytics in the business field, and the fact that employers are more and more looking for employees with skills related to data analytics, it is important for students to not only be successful in analytics courses, but to also have a positive experience. Upon developing and administering a survey to examine various components of students' perceptions going into analytics courses, we used factor analysis to identify the major factors contributing to these perceptions. The snapshot of the current landscape allows us to make decisions to incorporate changes to the course and help improve the experience for students going forward, with the goal that these decisions will help improve the perceptions of future students going in the course.

3 - Unsupervised Multi-Source Domain Adaptation in Cross-Machine Scenario with Mixed Heterogeneity

Yujun Yang, Yonsei University, Seoul, Korea, Republic of, Yongmin Kim, Hyunsoo Yoon

Rotating machinery is pivotal in industrial operations, where deep learning-based automated fault diagnosis offers significant advantages on system maintenance. With this premise cross-domain and cross-machine adaptation techniques are on a rapid research, where cross-domain adaptation deals with various load conditions and cross-machine adaptation aims to synchronize heterogeneous data collected from various machines which is more challenging. Multi-source based domain adaptation has garnered significant attention in research endeavors aimed at optimizing adaptation between source and target domains. However current multi-source based adaptation researches only deals with the scenario where multi sources and target shares the same heterogenity i.e in cross-domain scenario multi sources are given as various load condition and target as additional load condtion and in cross-machine scenario as various machines and additional machine. We claim that the intermediate scenario between aforementioned two can be more realistic where multi-source is given as various load and the objective target is given as additional machine. So we propose multi source unsupervised domain adaptaion method under mixed heterogeniety composed of multiple source under various load condition and target with additional machine with notable extent of domain gap.

4 - Analyzing a capacitated vehicle routing problem variant : Insights from approximations and mathematical analytics

Hyungjoo Cha, Korea University, Seoul, Korea, Republic of, Taesu Cheong

In this talk, we review the analysis of the capacitated vehicle routing problem in approximation and mathematical analytic point of views. Specifically, we formulate the CVRP-variant model for reverse logistics, where trucks are selectively dispatched to collect the parcels required. Apart from the mathematical formulation, we also execute the bound analysis using Beardwood-Halton-Hammersley theorem, which approximates the total tour length of a TSP tour with specific conditions. Moreover, with the modification of the well-known alpha-approximation algorithms, we provide a viable solution framework of the proposed algorithm. We review the solutions of the mathematical mode, approximation algorithms and the derived bounds correspondingly and derive the managerial insights. Our findings promise to inform decision-makers in logistics optimization, offering strategies to enhance efficiency and cost-effectiveness in transportation operations.

TA21

Summit - 341

Decision Analysis in Public Health and Biomedicine

Invited Session

Decision Analysis Society

Chair: Marie-Laure Charpignon, MIT, Cambridge, MA, United States

1 - Novel Digital Data Sources for Health Policy Research

Maimuna Majumder, Harvard Medical School & Boston Children's Hospital, Boston, MA, United States

In this talk, Prof. Majumder will discuss the utility of novel digital data sources for health policy research, with a focus on how sociopolitics —and politicians—have impacted our collective health during the COVID-19 pandemic. She will provide a brief overview of three of her lab's recent papers, covering the following interconnected topics: (1) Smartphone Mobility Data for National & State-Level Health Policy Evaluation; (2) Search Query & News Media Data for Monitoring Socioepidemic Misinformation; and (3) Social Media Data for Assessing Responses to Governmental Health Guidance. More specifically, she will focus on the impacts of (1) paid sick leave on compliance with non-pharmaceutical interventions in the US; (2) President Trump's remarks regarding the off-label use of disinfectants as a COVID-19 therapeutic; and (3) the CDC's mask guidance on the public's trust in American health agencies.

2 - Designing Trustable Risk-Based Intervention Strategies in an Infectious Disease Crisis

Andrew Perrault, The Ohio State University, Columbus, OH, United States

In an infectious disease crisis, planning interventions is a multi-objective problem that requires trading off among preventing the spread of disease, reducing the individual liberties of the uninfected, and financial and resource cost. In addition, intervention plans need to respect the information available to decision makers and be simple enough to be executed in a distributed fashion. I will present recent work where we take a cluster-based view that mirrors the information available to contact tracers. Using a simple agent-based model that can be updated as more detailed pathogen information becomes available, we develop an automated optimization approach leveraging deep and reinforcement learning that produces simple intervention strategies and achieves better objective values by about 25% compared to baselines.

3 - Monitoring Events in Large-Scale Public Health Data Streams

Ananya Joshi, Carnegie Mellon University, Pittsburgh, PA, United States

Traditional monitoring approaches were not designed to handle the volume and complexity of modern data. Their limitations are particularly apparent in the public health setting, where monitoring data volumes have increased exponentially and include incomplete, noisy, and non-stationary data subject to multiple forms of seasonality, censorship, and lags (revisions) that mask potential events like outbreaks and data quality issues. Our research addresses this gap by developing and implementing scalable monitoring methods tailored to handle the complexities of this modern monitoring setting. We present our novel approach, including new monitoring methods and frameworks that process and analyze vast amounts of health data in real-time to provide experts with actionable intelligence. Over the past 18 months, our approach has been deployed in a real-world public health setting and has demonstrated a 53x increase in monitoring efficiency. On average, over 200 events are identified weekly so data users can take informed decisions. This practical application opens new avenues for research in scalable monitoring techniques across domains.

4 - Regression discontinuity designs for evaluating AI/ML-enabled interventions in healthcare: Lessons learned from evaluations of two real-world deployments

Ben Marafino, Kaiser Permanente Division of Research, Oakland, CA, United States, Robert Gallo, Vincent Liu

While randomized trials are currently the gold standard for evaluating AI/ML-enabled interventions in health care, their implementation is often stymied by several factors. Integrating randomization into live clinical workflows frequently proves challenging, owing to complexities involved in both design and implementation. In addition, the oversight and compliance requirements associated with randomized trials, even “pragmatic” trials, can be burdensome. Given these limitations of randomized experiments in healthcare settings, regression discontinuity (RD) designs offer a credible alternative and turn out to be uniquely well-suited for analyses of the effectiveness of such AI/ML-enabled interventions. The RD design generally requires few assumptions and is capable of providing strong evidence without the need for randomization. Moreover, the RD design can be applied retrospectively, or it may form the bedrock of a planned analysis in instances where randomization proves infeasible. In this talk, I will briefly provide an overview of the unique features of the RD in the context of AI/ML-driven interventions, and present examples where it was successfully used to evaluate the effectiveness of large-scale AI/ML deployments at Kaiser Permanente and Stanford Health Care.

5 - New Approaches to Equitable Intervention Planning to Improve Engagement and Outcomes in a Digital Health Program

Yugang Jia, NA, Winchester, MA, United States

Background: Digital health programs provide individualized support to patients with chronic diseases and their effectiveness is measured by the extent to which patients achieve target individual clinical outcomes and the program’s ability to sustain patient engagement. However, patient dropout and inequitable intervention delivery strategies, which may unintentionally penalize certain patient subgroups, represent challenges to maximizing effectiveness. Therefore, methodologies that optimize the balance between success factors (achievement of target clinical outcomes and sustained engagement) equitably would be desirable, particularly when there are resource constraints.

Objective: Our objectives were to propose a model for digital health program resource management that accounts jointly for the interaction between individual clinical outcomes and patient engagement, ensures equitable allocation as well as allows for capacity planning, and conducts extensive simulations using publicly available data on type 2 diabetes, a chronic disease.

Methods: We propose a restless multiarmed bandit (RMAB) model to plan interventions that jointly optimize long-term engagement and individual clinical outcomes (in this case measured as the achievement of target healthy glucose levels). To mitigate the tendency of RMAB to achieve good aggregate performance by exacerbating disparities between groups, we propose new equitable objectives for RMAB and apply bilevel optimization algorithms to solve them. We formulated a model for the joint evolution of patient engagement and individual clinical outcome trajectory to capture the key dynamics of interest in digital chronic disease management programs.

TA22

Summit - 342

Decision Analysis for an Equitable Energy Future

Invited Session

Decision Analysis Society

Chair: Sheng Lun Cao, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Incorporating a Value of Lost Load Function In Equity-Centric Energy Systems Modeling

Sheng Lun Cao, Carnegie Mellon University, Pittsburgh, PA, United States

Energy system reliability is a key consideration in achieving energy equity for long-term grid-level systems planning, especially since long-duration outages due to increasing extreme weather impacts population groups differently. Incorporating equity-centric reliability considerations into energy systems modeling requires a re-evaluation of the traditional least cost optimization paradigm these models are typically based on. A key aspect of this incorporation is understanding people's preference heterogeneity and trade-offs for reliability. This relationship between people and reliability can be captured through the Value of Lost Load (VoLL), an estimated amount a customer is willing to pay to avoid a disruption to their electricity service. While typically a point estimate in $/kWh units, VoLL can be represented as a function relative to outage duration through the Customer Damage Function (CDF). We consider a) how a VoLL function can be used in reliability assessment of long-term grid-level systems planning, and 2) how the heterogeneous functional forms of VoLL can impact grid reliability at different spatial resolutions.

2 - Evaluating Equity Impacts of Vehicle-to-Grid Integration for the US Virgin Islands

Ogechi Vivian Nwadiaru, University of Massachusetts-Amherst, Amherst, MA, United States, Todd Levin, Erin Baker

This research integrates equity considerations into Vehicle to Grid (V2G) deployment, using the US Virgin Islands (USVI) as a case study. As distributed energy resources grow, vehicle-to-grid technologies can facilitate the transition to renewable energy by using electric vehicles as storage assets to mitigate potential renewable curtailment and offer resilience during periods of extreme weather events. Without deliberate attention to equity, V2G risks exacerbating disparities in access, affordability, and environmental justice. If done right, it has the potential to bring benefits to marginalized communities.

We investigate the grid and equity impacts of leveraging electrified public transit EV charging for renewable energy integration in a historically marginalized setting, focusing on the mid-transition. The study takes a two-prong approach, first, we optimize the charging schedule for public fleet EV deployment and V2G operational strategies and simulate the implications for system costs and reliability while also capturing key equity metrics. We then further iterate with community stakeholders to delve deep into unaddressed equity concerns.

To conduct this analysis we use the Argonne Low-carbon Electricity Analysis Framework (A-LEAF) as an integrated production cost model and reliability assessment toolkit. We demonstrate our approach through a case study application of the IEEE GMLC Reliability Test System (RTS-GMLC), which has been further calibrated to capture key parameters that characterize the USVI, such as the existing generation portfolio and demand profiles. Through this framework, we evaluate multiple V2G implementation and operational strategies and assess tradeoffs and complementariness across the stakeholder-derived metrics for localized reliability, cost, and mobility autonomy.

3 - Energy Justice Constrained Power System Dispatch and Planning

Juliette Ugirumurera, National Renewable Energy Laboratory, Golden, CO, United States

Power plants are one of the main sources of air pollution. Our current grid modeling capabilities are unable to represent air pollution and health impacts from the electricity industry in the optimization. Therefore, we propose a novel approach to link a power system model (Sienna) with a high-fidelity air quality model (known as SCICHEM) and human health impacts modeling tool (known as BenMAP) to assess the impacts of operating power plants on health and damage costs. This linkage will help us develop a module to evaluate the environmental justice impacts on different demographic groups, especially vulnerable communities. It will also help inform better decisions in capacity expansion models for future systems to avoid unfairly burdening certain communities. A case study will demonstrate modeling capabilities and generate example analysis for regions with at-risk communities. This new capability can be used by system operators, researchers, and other entities to better plan and operate energy systems that are just for all and to understand the implications of our current operating power system on specific communities.

4 - Advancing equitable household and community benefits through energy storage

Jessica Kerby, PNNL, Salem, OR, United States, Bethel Tarekegne

As energy prices rise and climate change brings more extreme and frequent days of heating and cooling, households must allocate more of their income to energy bills, increasing their energy burden. Many strategies are employed to alleviate high energy burden, such as weatherization, energy efficiency, and energy storage and rooftop solar, though the benefits of each scale based on factors such as climate, housing characteristics, and energy behaviors. In addition, federal and state decarbonization goals have led to numerous financial incentives and policies designed to increase access and adoption of renewable energy systems. In combination with the declining cost of both solar photovoltaic and battery energy storage systems and rising electric utility rates, residential renewable adoption has become more favorable than ever. However, not all states provide the same opportunity for cost recovery, and the complicated and changing policy and utility landscape can make it difficult for households to make an informed decision on whether to install a renewable system. This work explores the idea of providing a targeted approach to energy burden reduction, based on household conditions, climate, and policy. For households considering renewable adoption, this work also discuss the relevant factors that influence renewable system performance and payback as a guide to help navigate the policy, utility rate structures, system, and financial considerations required for rooftop solar and energy storage.

5 - Locational Marginal Burden: Equity embedded in optimal power flow

Amanda West, Georgia Institute of Technology, Atlanta, GA, United States, Samuel Talkington, Rabab Haider

The fair distribution of benefits in electric power systems is a pertinent energy policymaking problem; however, energy justice and equity efforts cannot be easily quantified in existing power system engineering studies. To incorporate justice and equity into operations and planning problems, we propose a metric termed *locational marginal burden* (LMB). The LMB metric provides an interface between energy burden -- well-studied in energy pricing equity -- and the optimal power flow problem -- used ubiquitously in electricity market clearing and resource dispatch -- by leveraging tools from differentiable optimization. In this work we present the LMB metric which describes the marginal impact of nodal demand in the transmission system to the energy burden experienced by customers, similar to traditional Locational Marginal Prices (LMP) used in transmission wholesale electricity markets. In deriving the LMB metric, we can assess the impact to energy burden of a customer by an increase in both local demand (at the same transmission node) as well as network-wide effects (at other transmission nodes). This presentation introduces the LMB metric, preliminary investigation of LMB-integrated planning decisions, and future directions of research in tariff design and integrated resource planning.

TA23

Summit - 343

Digital Content and Platform Strategies

Contributed Session

Chair: Wei Lu, University of Science and Technology of China, 96 Jinzhai Rd, Hefei, 230026, China, People's Republic of

1 - Subscription Mix: Menu Pricing in Two-Sided Media Platform Competition

Yunhyoung Kim, University of Minnesota Twin Cities, Minneapolis, MN, United States

This article studies the impact of subscription mix, "consumers' multi-homing behaviors based on video service platforms' menu pricing strategies", in two-sided media platform competition. Recently, there are huge portions of consumers who subscribe to more than one video streaming service. Platforms like Netflix and YouTube have introduced varied subscription models—including premium, ad-supported, and free access—to not only increase their subscriber base but also to appeal to a broader range of advertisers. The study questions the efficacy of such strategies, considering the unpredictability of consumer subscription combinations and the potential decline in advertising revenue if platforms fail to target unique consumer segments effectively. Given these dynamics, two different media platforms are considered: an advertising-focused media platform and a subscription-focused media platform. An analytical model, grounded in two-sided platform competition literature, is employed to explore these phenomena.

Key findings reveal that consumers' mix of subscriptions encourage platforms to adopt asymmetric menu pricing strategies, allowing them to mitigate intense competition on both the consumer and advertiser sides. These strategies are found to be largely dependent on the marginal utility derived from multi-homing. There is an identified optimal level of marginal utility increase from multi-homing that enables platforms to diversify subscription plans and maintain a competitive edge, resulting in mutual benefits for all involved up to certain thresholds. The advantageous effects diminish as the marginal value from multi-homing changes, presenting a critical balance that platforms must navigate.

2 - Video Views or View Time: Managing Platform Compensation for Content Creators

Xu Zhang, Fudan University, Shanghai, China, People's Republic of, Hongbo Zha, Yinuo Tang, Yifan Dou

For content platforms, compensating content creators for their contributions is pivotal to the success of a thriving ecosystem. We use video content platforms such as YouTube, TikTok, and Bilibili as examples in this paper. Traditionally, compensation has been based on the count of Video Views (VV), which is directly linked to the platform's revenue from advertising services. However, to incentivize long-video creators, this paper evaluates a new compensation model based on View Time (VT), which is emerging in practice. The existing literature has not sufficiently examined how incentive compensation schemes should accommodate view time, which represents a distinct dynamic within the creator economy. We develop a game-theoretical model to investigate whether transitioning from VV to VT is beneficial for the platform, creators, and viewers.

The VV-VT transition helps resolve the traditional conflict of compensating long-video creators, specifically sharing less revenue in the time-constrained viewers' market while incentivizing creators. Additionally, the VV-VT transition can enhance the relative attractiveness of long videos, which previously faced a disadvantage compared to short videos, thereby altering the competitive structure. This enhancement is, however, negatively affected by the proportion of time-constrained viewers. While insufficient or moderate enhancement can intensify the competition among creators, excessive enhancement will conversely mitigate the competition. Therefore, a higher proportion of long-video creators and a higher proportion of time-constrained viewers increase the profitability of the VV-VT transition. Furthermore, with a significantly large proportion of time-constrained viewers, we find that the transition can lead to a Pareto improvement.

3 - Content Creation, Competition and Selective Promotion of Apps

Aditya Karanam, National University of Singapore, Singapore, Singapore, Ashish Agarwal, Anitesh Barua

It is well documented in the literature that selective promotion of mobile applications in the app store has cascading effects on the demand and design of a range of non-promoted applications. However, the nature of influcence of these spillover effects on the category's overall demand and competition is not clearly known, especially in the context of app econonomy. By utilizing iOS App Store data spanning the years 2017 to 2022, this research seeks to understand the impact of these spillover effects on the overall design and competition among applications in a given category. Our results indicate that non-promoted apps within a category become increasingly similar to the promoted apps. We also observe that while the overall demand of a category increases, this increase is majorly concentrated among a few apps. These findings underscore the significant role played the platform promotions on the design and competition of the overall ecosystem.

4 - Retailer's Information Sharing and Manufacturer's Channel Expansion in The Live-Streaming E-commerce Era

Wei Lu, University of Science and Technology of China, Hefei, China, People's Republic of, Jie Wu, Xiang Ji

Numerous manufacturers started to embrace live-streaming selling channels in addition to their pre-existing retail channels during the outbreak of COVID-19. Our work investigates a retaile's optimal strategy for sharing demand information with a manufacturer who might collaborate with a streamer to build a live-streaming selling channel. The results point out that the manufacturer's live-streaming selling channel expansion via an influencer does not necessarily harm the retailer because the retailer can free ride on the market expansion due to the social influence of the streamer. We also provide a rationale for the widespread voluntary information sharing observed in the era of live-streaming selling. The retailer can discourage or encourage the manufacturer to establish a live-streaming channel by sharing the demand information, which depends on the streamer's social influence. In addition, potential changes to information sharing policy result in several unexpected profit implications for the manufacturer, whose profit exhibits a nonmonotonic relationship to the streamer's social influence and the channel expansion cost. In other words, our results counterintuitively show that the manufacturer may suffer from cooperating with a highly influential streamer in a live-streaming channel but benefit from choosing a less influential one.

TA24

Summit - 344

Applications and Implications of Large Language Models

Contributed Session

Chair: Senyao Gu, Tsinghua University, Beijing, N/A

1 - *Examining the Impact of Reliance on LLMs on Perceived Creativity*

Jiaoping Chen, Michigan State University, East lansing, MI, United States, Laura Brandimarte, Anjana Susarla

Large Language Models (LLMs) have been increasingly integrated into a myriad of open-ended tasks, facilitating human endeavors in generating creative outputs ranging from product ideation to digital artwork. Such novel capabilities of LLMs have ushered in a new era of human-AI collaboration. Limited studies focus on the impact of reliance on AI's advice and the generative capabilities of LLMs. In this study, we address this gap using creative writing tasks. We first employ a structured experimental framework by focusing on creative writing tasks under two distinct conditions: low vs high LLM randomness. We then assess two outcome variables: perceived creativity, and likability of these narratives by conducting an online survey. This study finds that the effect of writers' reliance on AI advice on perceived creativity varies with the level of generative capabilities inherent in the LLMs (high vs. low randomness). Specifically, in contexts where LLMs generate less diverse suggestions, a higher acceptance of AI suggestions is associated with decreased perceived creativity. In contrast, when LLMs provide more diverse advice, the increasing reliance on LLMs are associated with increased creativity of work. Lastly, we further explore how the characteristics of raters could influence their perceptions of creativity.

2 - Distillation of knowledge and opinion of LLMs in an opinion dynamic framework

Lan Hoang, IBM, Warrington, United Kingdom, Michail Smyrnakis, Carol Mak

Large language models (LLMs) have emerged as a group of powerful, versatile models that can assist humans in multiple tasks. To date the focus has largely been on single LLM or a mixture of LLMs to conduct tasks. Yet, there is an implication for knowledge and bias distillation arising with LLM interactions, such as in via fine-tuning or interactions. In this work, we aim to review the current approaches of social interactions of LLMs and propose a framework to investigate the interactions using agent-based modelling. Specifically we will use the opinion dynamic experiments and equilibrium analysis to enable the study of knowledge and bias propagation. This cascading through the networks can lead to LLM clusterings and group dynamics that differ from initial training objectives. This direction will have applications in the area of AI ethics and in the real world application of LLMs in the multi-agent setting.

3 - Patterns for Co-opting Large Language Models and Knowledge Bases to Enable Smarter Decisions

Dinesh Verma, IBM Research, Yorktown Heights, NY, United States, Pawan Chowdhary, David Beymer, Swanand Kadhe, Jon Lenchner, Shiqiang Wang

While Large Language Models (LLMs) have proven themselves useful in many contexts, they also have drawbacks including hallucinations and non-repeatability. In many cases, decision making based on knowledge bases may be better from a reliability, explainability or repeatability perspective, but suffers from the high overhead required for maintaining and updating the knowledge base. In order to make smart dynamic decisions, the best approach may often consist of a hybrid approach that combines the strengths of both LLMs and knowledge bases. This combination can be done in many different ways, and results in a few distinct patterns for joint use of LLMs and Knowledge bases. In this paper, we will present and examine some of the common patterns for joint use of LLMs and knowledge bases and compare their strengths and weaknesses. Specifically, we consider the following patterns: (a) using the knowledge base as a backup mechanism for a LLM (b) using a LLM to augment a knowledge base (c) using a knowledge base to fact-check a LLM and (d) using a knowledge base to optimize inference based on LLMs, among others. We will also show the effectiveness of the patterns in solving a problem such as secure search on the Internet.

4 - Integrating Large Language Models into Production Planning: A Framework for Simulation-Based Optimization

Lennard Heuer, Korea Advanced Institute of Science and Technology, Daejeon, Korea, Republic of

Creating optimization models (OMs) for production planning within a simulation environment that closely replicates a real-world manufacturing setting is challenging, as these OMs often fail to encapsulate the complexity and dynamic relationships inherent in such simulations. However, OMs offer interpretability beyond numerical results alone, revealing crucial heuristics and contextual relationships essential for strategic decision-making. Large language models (LLMs) have recently surged in popularity and can now be utilized in diverse tasks, including code generation. Leveraging the capabilities of LLMs, this study presents a framework for creating customized, LLM-generated production planning OMs, using detailed simulation feedback for OM refinement. Initially, an LLM processes user input data to select or combine linear programming OMs from a preselected pool. Next, using an iterative approach, an LLM extends the OM, potentially transforming it into more complex forms such as mixed-integer nonlinear programming. The LLM thereby modifies the objective function and adds constraints focused on production, sales, inventory, and workforce strategies. Simulation feedback is used to evaluate the overall performance of the production plan derived from an OM and to enable an LLM to identify specific weaknesses detected in the simulation feedback. This allows for intelligent decisions on further LLM-generated alterations of the OM. The above-described method is embedded as a mutation in a genetic algorithm, which also utilizes population initialization, selection, and crossover to enable a global search for an optimal OM. Furthermore, OM parameters newly added through the genetic algorithm are fine-tuned for further optimized production plan performance in simulation.

5 - Utilizing Large Language Models to Elicit Human Thought Processes in Decisions under Risk

Senyao Gu, Tsinghua University, Beijing, China, People's Republic of, Chen Wang

This study performs cumulative reasoning on a large language model (LLM) to reveal the underlying thought processes based on a mixture of decision theory models. The LLM is specifically tuned to mimic human medical decisions in a large dataset of electronic health records. We first prompt the LLM to qualitatively classify if the basic arithmetic operations (including monotonicity, convexity, continuity, and isomorphism) are employed to reach the treatment decision. We then fit a mixture of decision theory models (MoT) based on the classification results to depict the thought processes quantitatively. This study extends recent advances in using deep learning methods to describe human decisions under risk, with simple lottery experimental data, to real-world decision contexts involving multi-modal data (qualitative, quantitative, texts, etc.)

TA25

Summit - 345

Gurobi/Stukent

Invited Session

Technology Showcase

1 - Practical Guidelines for Model Improvement and Reformulation

Rodrigo Fuentes, Gurobi Optimization, Plymouth, MA, United States

In this showcase, we will share insights and lessons learned from helping Gurobi customers from a wide range of industries adjust their optimization models to improve solver performance and numerical behavior. We will look at the challenges that we see most often in LP, MIP and MINLP models, and discuss our approach and typical recommendations to help address them. We will also consider some well-known modeling “rules of thumb” and discuss how applicable they are in 2024.

2 - How Virtual Work-integrated Learning can Help Educators Overcome Today’s Instructional Challenges

Scott Carr, Stukent, Idaho Falls, ID, United States

A Stukent Simternship is a career-relevant, work-integrated learning experience that helps students connect classroom concepts to real-world tasks. In a Simternship, students become marketing managers, PR officers, entrepreneurs, SEO specialists, accountants, and more. They interact with simulated supervisors and coworkers, perform realistic tasks, and build confidence in the safety of a simulated environment.

Simternships allow students to apply new knowledge to professional problems. This level of immersion helps students encode the information they are learning in class, which allows them to transfer new concepts to real-world scenarios. All Simternships feature auto-grading and LMS integration to take the hassle out of hands-on education.

TA26

Summit - 346

Explainable Machine Learning for Knowledge Discovery

Invited Session

Quality, Statistics and Reliability

Chair: Yuhao Zhong, Texas A&M University, College Station, TX, 77840, United States

Co-Chair: Satish Bukkapatnam, Texas A&M University, College Station, TX, United States

Co-Chair: Ashif Iquebal, Arizona State University, Tempe, AZ, United States

1 - Quantifying Causal Influences with Graphical Causal Models for Root Cause Analysis and Beyond

Patrick Bloebaum, AWS, Santa Clara, CA, United States

Explainability has become a critical challenge for businesses seeking to make informed decisions and drive continuous improvement. In this talk, we present domain-agnostic approaches that utilize graphical causal models to represent causal relationships within a system, enabling the quantification of direct and indirect causal influences of system components on key metrics. This allows us to move beyond mere statistical associations to gain a deeper understanding of a system. We particularly focus on novel concepts for root cause analysis that allow the identification of the underlying causes of anomalies and distribution changes by modeling the causal mechanisms that generated the data. Furthermore, we illustrate how this causal approach can be used to explain the factors influencing specific key metrics in a system, which is crucial for any critical decision-making process. The presented concepts and algorithms in the talk are readily available as open-source solutions and can be integrated into various business processes to enhance explainability, optimize systems, and drive data-driven decision-making.

2 - Brple: Bayesian Regularized Post-Hoc Local Explanations

Yuhao Zhong, Texas A&M University, College Station, TX, United States, Anirban Bhattacharya, Satish Bukkapatnam

Existing studies that focus on explaining black-box machine learning models either overlook the uncertainty in generated explanations or quantify this uncertainty inadequately. To address this gap, we propose BRPLE to explain machine learning models and obtain the distribution of feature importance using Bayesian ridge regression. We present the mathematical framework of our Bayesian approach and discuss theoretical outcomes, including the significance of the ridge parameter. Our methodology was validated through case studies on both numerical simulations and benchmark datasets. Compared to state-of-the-art methods, BRPLE offers superior accuracy in uncertainty quantification, specifically in deriving the posterior distributions and rankings of feature importance.

3 - Explainable Machine Learning for Knowledge Discovery in Advanced Manufacturing Systems

Joseph Cohen, University of Michigan, Ann Arbor, MI, United States, Eunshin Byon, Xun Huan

In recent years, machine learning has become popular for optimizing the operations of advanced manufacturing systems. These data-driven algorithms have enabled key performance indicator improvements such as reducing machine downtime, improving quality, yield, and sustainability. However, as advanced data analytics and digital twin platforms mature in the age of generative AI, there is an unmet need to critically evaluate model outputs to gauge trustworthiness. This presentation will focus on our efforts in advancing explainable anomaly detection, fault diagnosis, and prognosis of high-dimensional applications in semiconductor and aerospace manufacturing systems, as well as renewable power systems. The talk will detail an approach for semi-supervised clustering based on Shapley value analysis, with the added capability of accurately describing clusters with measurable inputs. In addition, our efforts in explaining model outputs via counterfactual explanations will be discussed. Lastly, we will describe how these techniques provide actionable insights for hypothesis generation, knowledge discovery, and synergize well with other AI application areas such as active learning and reinforcement learning.

4 - Multi-perspective Explainable Artificial Intelligence

Andrew Kusiak, University of Iowa, Iowa City, IA, United States

Explainable artificial intelligence (XAI) solutions provide insights into digital models and increase understanding and confidence in the generated results. As data-derived models become more complex, wider is scope, and cover multiple domains, interpretation of the model produced decisions becomes challenging. Such decision-making scenarios call for multi-domain expertise to design XAI systems for the deployment of models and interpretation of the predicted outcomes. An XAI solution for complex and multi-domain applications is discussed. The solution includes different modes of operations and supports user preferences.

5 - A Stochastic Method for Physics Discovery via Sparse Identification of Nonlinear Dynamics

Ridwan Olabiyi, Arizona State university, Tempe, AZ, United States, Han Hu, Ashif Iquebal

This paper introduces a novel stochastic method that leverages the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm and consistent Bayesian inference to derive physics models from data. This approach not only identifies crucial dynamical terms in the system governing equations but also determines the probabilistic distributions of their coefficients, thereby producing a detailed model that accurately reflects the underlying physical processes. Our method is particularly suited for inferring the unknown governing physics of complex systems characterized by nonlinear dynamics with stochastic parameters. The proposed method not only provides a robust framework for discovering governing equation of such complex system from data but also allows for uncertainty quantification of the discovered models under varying conditions, making it an invaluable tool in machine learning frameworks for physics discovery. We evaluate the efficacy of our proposed methodology against conventional methods through benchmark datasets using simulated data from the Lotka-Volterra model, the Hudson Bay (lynx-hare) dataset, and three diverse infiltration datasets. The results shows that our method outperforms existing approaches based on standard metrics, confirming its superiority in modeling complex dynamical systems.

TA27

Summit - 347

Rethinking Human-AI Interactions

Invited Session

New Product Development

Chair: Evgeny Kagan, Johns Hopkins Carey Business School, Baltimore, MD, United States

1 - Learning in Human&Mdash;AI Teams

Zhaohui Jiang, Carnegie Mellon University, Pittsburgh, PA, United States, Param Singh, Linda Argote

Although rapid developments in algorithms and artificial intelligence (AI) are occurring, these AI systems often do not operate fully autonomously but rather require humans and AIs to jointly make decisions. Because humans and AI systems have complementary skills and information, their joint decision making could be better than human or AI decision making alone. We refer to these collaborative decision-making units as *human—AI teams* and seek to understand the human learning process in human-AI teams, including: (1) Can humans learn how to collaborate with AI through repeated practice? (2) What factors affect this learning process? Does the interpretability of the AI collaborator impact learning, and if so, what is the optimal level of interpretability? (3) Does human independent decision-making ability matter, and how would it impact the answers to the questions above? Findings from our studies can offer policy and managerial insights (e.g., designing more effective training programs) to facilitate modern decision making with human-AI teams.

2 - Effect of Task and Temporal Attributes on Customers&Rsquo; Decisions to use Self-Service Technologies

Yinghao Zhang, University of Cincinnati, Cincinnati, OH, United States, Chia-Chun Yang

Self-service technologies (SSTs) are being increasingly implemented across various service industries, including financial, retail, and healthcare. Early research highlighted the benefits of SSTs for service organizations, emphasizing cost reduction and productivity enhancement. Recent studies have shifted their focus toward understanding the factors that motivate customers' decisions to use SSTs. These studies have primarily concentrated on the factors of customer characteristics (e.g., personality, technology attitude, self-efficacy, and prior experience) and system design (e.g., convenience, ease of use, controllability, flexibility, supportiveness, and enjoyment). However, there is still a lack of knowledge regarding situational factors that can influence customers' decisions. Hence, our study focuses on two such factors: Task Complexity and Time Pressure. This leads to our research question: How do task complexity and time pressure influence customers' preferences of using self-service technologies? To answer this question, we designed and conducted online experiments to observe people's SSTs preferences under a simulated grocery checkout environment. This research method is novel compared to the previous literature, which typically relies on scenario-based surveys to examine people's general perceptions and choices. Our study contributes to the existing literature by examining situational factors as determinants of customers' decisions to use service technology; where knowing how such factors affect customers' decisions is insightful for organizations to make better decisions about investing and allocating self-service technologies to improve operational and financial performances.

3 - Automation and Augmentation: Roles of AI in Collaborated Decision Making

Andreas Fuegener, University of Cologne, Köln, Germany, Dominik Walzner, Alok Gupta

Artificial intelligence (AI) will have a growing influence in the future of work. Human decision-makers may see significant changes in their day-to-day work as collaboration between humans and AI will become commonplace. We explore the application of AI for automation (i.e., AI performing tasks independently) and for augmentation (i.e., AI advising humans) in collaborative environments. Using an analytical model, we show that whether AI should be used for automation or for augmentation depends on different types of human-AI complementarity: The share of automation increases with higher levels of between-task complementarity, which can arise due to task-level performance differences between humans and AI. In contrast, the share of augmentation increases with higher levels of within-task complementarity, which arise due to task-based interaction between humans and AI. We include both AI roles in a task allocation framework, where an AI and humans work on a set of classification tasks to optimize performance with a given level of available human resources. We validate our framework with an empirical study based on experimental data in which humans had to classify images with and without AI support. When between-task and within-task complementarity exist, we see an interesting distribution of work pattern for optimal work configuration: AI automates relatively easy tasks, augments humans on tasks with similar human and AI performance, and humans work without AI on relatively difficult tasks. Our work provides several contributions to theory and practice and our task allocation framework showcases potential job designs in the future of work.

4 - Biases of Humans, of AI, and of Humans with AI

Tracy Jenkin, Queen's University, Kingston, ON, Canada, Yang Chen, Sam Kirshner, Anton Ovchinnikov, Meena Andiappan

Since the release of ChatGPT in November 2022, research on decision-making biases in large language models (LLMs) has rapidly proliferated. LLMs are mathematical models that represent the current frontier of Artificial Intelligence (AI), but they are trained on human data, and a consistent observation from this research is that LLMs may mirror human biases, remain unbiased, or can even display biases distinct from those of humans. Our research advances the literature by examining whether individuals change their inherent biases when interacting with AI that might be biased similarly or differently from humans. To study this question, we design a novel interactive AI experimental design paradigm and apply it to three representative biases from the literature. We show that AI biases indeed impact human decision-making and discover a new form of nudging via system prompts. Finally, we propose a new take on algorithm aversion by showing that how humans interact with AI matters for improving decision-making.

TA28

Summit - 348

Smart Manufacturing Systems

Contributed Session

Chair: Thomas Meiren, Fraunhofer IAO, Nobelstr. 12, Stuttgart, 70569, Germany

1 - Time-Series Classification in Smart Manufacturing Systems

Mojtaba Askarzadeh Farahani, West Virginia University, Morgantown, WV, United States, Matt McCormick, Ramy Harik, Thorsten Wuest

The shift toward smart manufacturing marks a new era dominated by digital technology, significantly enhancing data collection capabilities through advancements in sensing technologies. Central to these smart manufacturing systems (SMS) is the utilization of time-series data, making Time-Series Classification (TSC) a critical focus area. Our study aims to bridge a gap in current research by conducting a detailed experimental evaluation of modern machine learning (ML) and deep learning (DL) algorithms tailored for TSC tasks within manufacturing and industrial settings. Initially, we identified 92 algorithms from the realms of TSC and manufacturing. We methodologically selected 36 of these algorithms as the most indicative of state-of-the-art performance. These selected algorithms were then evaluated using 22 manufacturing datasets, chosen to represent different challenges in the industry. The evaluation revealed that ResNet, DrCIF, InceptionTime, and ARSENAL are the standout algorithms, all demonstrating superior accuracy across the datasets. Additionally, the study highlights the capabilities of LSTM, BiLSTM, and TS-LSTM algorithms, which excel in processing sequential data through recurrent neural network architectures. This comprehensive testing and evaluation provide practitioners with clear insights into the effectiveness of various ML and DL models, guiding better decision-making in the application of TSC techniques in smart manufacturing.

2 - Framework and Future Trends for Industry 4.0 Technologies

Arvin Shadravan, Texas A&M University, College Station, TX, United States, Hamid R. Parsaei

In the digital age, manufacturing has undergone a profound transformation due to advancements in data capture systems, information technology, and network technologies. This global shift towards digitalization introduces challenges stemming from the rapid pace of technological evolution. Advanced manufacturing strategies, notably Industry 4.0, have emerged in response, aiming to achieve "smart manufacturing". Today, smart manufacturing emphasizes technologies such as cyber-physical systems, digital twins, cloud computing, the internet of things, and big data analytics, marking a shift from knowledge-based to data-driven processes. Despite the challenges, many companies have adopted reference architectures to guide system interoperability and structure. However, the suitability of these architectures varies based on specific use cases and lacks a comprehensive understanding of existing reference architecture. This study reviews and analyzes current reference architectures for Industry 4.0, assessing their effectiveness in supporting smart manufacturing processes. It also highlights the urgent steps needed for improvement and underscores the importance of continued investment in reference architectures to secure the future of Industry 4.0. The pervasive impact of digital technologies across various industries necessitates comprehensive digital transformations involving the revamping of organizational structures, business models, and processes. Evaluating an organization's digital maturity is crucial for crafting effective digitalization strategies tailored to specific needs, unlocking potential throughout the transformation process.

3 - The Role of Smart Services in European Mechanical Engineering Companies

Thomas Meiren, Fraunhofer IAO, Stuttgart, Germany

The mechanical engineering industry has invested heavily in digitalization in recent years. This applies not only to manufacturing, but also to its service business. In particular, new digitally-enabled services are on the rise, using the growing volume of data to create new personalized solutions for customers. Examples of such 'smart services' include condition monitoring, predictive maintenance and smart training. The results of a study of the European engineering industry will be presented at the conference. In particular, the study looked at the role that smart services currently play in the service portfolio, the digital technologies on which they are based, and how they are perceived to create value for customers. The study also analyzed the differences between small and medium-sized enterprises and large companies.

4 - Generative Manufacturing Systems Using Diffusion Models and ChatGPT

Xingyu Li, Purdue University, West Lafayette, IN, United States

In this study, we introduce Generative Manufacturing Systems (GMS) as a novel approach to effectively manage and coordinate autonomous manufacturing assets, thereby enhancing their responsiveness and flexibility to address a wide array of production objectives and human preferences. GMS envisions the seamless integration of stationary machinery, autonomous assets, and diverse human workforces within futuristic manufacturing systems. In light of the growing autonomy and mobility of assets, we propose that autonomous assets and humans can dynamically relocate and self-organize across various workstations to optimize manufacturing operations and material flows.

To address the decision complexity associated with the growing number of autonomous assets and to align with human desires, our research focuses on developing novel generative AI methods based on diffusion models and ChatGPT. These methods aim to provide specialized decision support in critical manufacturing decisions such as capacity planning, layout design, and operation scheduling, responding to human textual or voice inquiries. For example, queries such as "I require a minimum production rate of 180 parts per hour" or "I specialize in collaborating with robots in Task X" are accommodated by the proposed generative AI models, offering multiple quality decisions for human selection.

We have implemented GMS in an industrial use case for part processing. Our empirical findings demonstrate GMS's significant potential in enhancing system resilience and responsiveness to uncertainties, with decision times reduced from seconds to milliseconds. The study highlights the inherent creativity and diversity in the generated solutions by GMS, facilitating human-robot collaborations and human-centric decision-making through continuous refinement.

TA29

Summit - 420

Topics in Integer Programming and Combinatorial Optimization 1

Invited Session

OPT: Integer and Discrete Optimization

Chair: Joseph Paat, University of British Columbia, Sauder School of Business, Vancouver, BC, Canada

Co-Chair: Luze Xu, University of California, Davis, Davis, CA, 95616, United States

1 - On the Congruency-Constrained Matroid Base

Chao Xu, University of Electronic Science and Technology of China, Chengdu, China, People's Republic of, Siyue Liu

Consider a matroid where all elements are labeled with an element in Z. We are interested in finding a base where the sum of the labels is congruent to $g \pmod{m}$. We show that this problem can be solved in $\tilde{O}(2^{4m} n r^{5/6})$ time for a matroid with n elements and rank r, when m is either the product of two primes or a prime power. The algorithm can be generalized to all moduli and, in fact, to all abelian groups if a classic additive combinatorics conjecture by Schrijver and Seymour holds true. We also discuss the optimization version of the problem.

2 - Two Relaxations of Woodall's Conjecture

Siyue Liu, Carnegie Mellon University, Pittsburgh, PA, United States, Ahmad Abdi, Gerard Cornuejols, R Ravi, Olha Silina

In a digraph, a dicut is a cut where all the arcs cross in one direction. A dijoin is a subset of arcs that intersects each dicut. It is proven that the minimum size of a dicut, say k, equals to the maximum fractional packing of dijoins. Woodall conjectured in 1976 that the maximum integral packing also equals to k. In other words, it is conjectured that the minimum size of a dicut equals to the maximum number of disjoint dijoins. We prove two relaxations of Woodall's conjecture. First, we show that Woodall's conjecture is true up to a constant factor. More precisely, there exist k/6 disjoint dijoins. Second, we show that if we relax the nonnegativity of the packing coefficients, there exists an integral packing of k dijoins.

3 - A characterization of unimodularity of hypergraphs with disjoint hyperedges

Marco Caoduro, University of British Columbia, Sauder School of Business, Vancouver, BC, Canada, Joseph Paat

We are motivated by integer linear programs (ILPs) with bounded determinants in the constraint matrix. A canonical example is the maximum stable set problem in a graph, which can be modelled as an ILP where the largest determinant is exponential in the graph's odd cycle packing number. This follows from a characterization of determinants of incidence matrices by Grossman et al. During the talk, we explore a possible extension of Grossman et al.'s characterization to hypergraphs.

4 - Extended formulations for some strictly Delta-modular IPs

Joseph Paat, University of British Columbia, Sauder School of Business, Vancouver, BC, Canada, Luze Xu, Zach Walsh

Conforti et al. give a compact extended formulation for a particular class of bimodular-constrained ILPs, namely those that model the stable set polytope of a graph with odd cycle packing number 1. We extend their techniques to design compact extended formulations for the integer hull of translated polyhedral cones whose constraint matrix is strictly Delta-modular and has rows that represent a cographic matroid. This is joint work with Zach Walsh and Luze Xu.

5 - Proximity bounds and structural properties for Δ-modular integer programming

Luze Xu, University of California, Davis, Davis, CA, United States

The maximum absolute subdeterminant of the coefficient matrix is known to be an important parameter for measuring the complexity of integer programming, where an open conjecture is whether integer programs can be solved in polynomial time when the maximum subdeterminant is fixed. We call an integer matrix Δ-modular if its maximum absolute subdeterminant is Δ. In light of this conjecture, natural questions are to understand structural properties of Δ-modular matrices and to investigate proximity bounds: how large can the distance be between an optimal vertex solution of the corresponding linear program relaxation and the closest integer optimal solution in the integer linear program?

We provide a partial list of matrices that cannot exist in a Δ-modular matrix. Using Wolsey's b-hull results, we propose a method to determine the proximity bound for a fixed matrix with parametric right hand sides and objective functions. Furthermore, by adapting an algorithm by Averkov and Schymura for classifying rank-m Δ-modular matrices up to unimodular transformation, we calculate sharp proximity bounds for small m and Δ using SageMath.

Based on joint work with Matthias Köppe, Moises Reyes Rivas, Joseph Paat, Ingo Stallknecht, and Zach Walsh.

TA30

Summit - 421

Optimization Software Session III

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelmann, Arizona State University, Tempe, AZ, United States

1 - Cupdlp: a Gpu Implementation of Restarted Primal-Dual Hybrid Gradient for Linear Programming

Haihao Lu, The University of Chicago, Chicago, IL, United States, Jinwen Yang

In this paper, we provide an affirmative answer to the long-standing question: Are GPUs useful in solving linear programming? We present cuPDLP.jl, a GPU implementation of restarted primal-dual hybrid gradient (PDHG) for solving linear programming (LP). We show that this prototype implementation in Julia has comparable numerical performance on standard LP benchmark sets as Gurobi, a highly optimized implementation of the simplex and interiorpoint methods. Furthermore, we present the superior performance of cuPDLP.jl with its CPU counterpart. This demonstrates the power of using GPUs in the optimization solvers.

2 - The Sas Optimization Solvers: Recent Improvements and Future Developments

Bochuan Lyu, SAS Institute Inc., Houston, TX, United States, Philipp Christophel, Laszlo Ladanyi, Menal Guzelsoy, Amar Narisetty, Rob Pratt

In this talk we present the recent improvements to the SAS linear and mixed integer linear optimization solvers and discuss future developments. Topics include Python interfaces to SAS Optimization, automated benders decomposition, presolve improvements, and cutting plane selection.

3 - Progress on Highly Parallel Ensemble Solvers

Yuji Shinano, Zuse Institute Berlin, Berlin, Germany

The Ubiquity Generator (UG) Framework was initially developed to parallelize advanced branch-and-bound based solvers. With the introduction of version 1.0, UG transitioned into a high-level task parallelization framework, enabling the parallelization of a broad range of state-of-the-art solvers. This framework facilitates the integration of multiple algorithmic implementations into a parallel solver. This presentation will highlight several success stories where UG has been utilized to solve previously intractable instances of Mixed Integer Programming, Steiner Tree, Quadratic Assignment, and Quadratic Unconstrained Binary Problems, as well as to set new benchmarks for Shortest Vector Problem instances. Finally, the most recent developments in the UG Framework will be discussed.

4 - Recent Progress in the Global Optimization of Nlps and Minlps with Baron

Nikolaos Sahinidis, Georgia Institute of Technology, Atlanta, GA, United States

We review historical progress and recent developments toward the global optimization of nonlinear and mixed-integer nonlinear programming problems with BARON. We discuss relaxation construction, preprocessing, branching, and tree management techniques. In addition to widely accepted benchmarks, we present results for a large test set.

TA31

Summit - 422

Recent Advances in First and Second Order Methods for Nonlinear Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Ya-Chi Chu, Stanford University, Stanford, CA, 94305, United States

1 - Exploiting Negative Curvature in Conjunction with Adaptive Sampling

Albert Berahas, University of Michigan, Ann Arbor, MI, United States, Raghu Bollapragada, Wanping Dong

In this talk, we present methods that exploit negative curvature for stochastic unconstrained optimization problem.

2 - Second-Order Method for Solving Monotone Variational Inequalities: An Appoximation-Based Regularized Extra-Gradient Approach

Kevin Huang, National Taiwan University, Taipei, Taiwan, Shuzhong Zhang

In this paper, we propose a general extra-gradient scheme for solving monotone variational inequalities (VI), referred to as Approximation-based Regularized Extra-gradient method (ARE). The first step of ARE solves a VI subproblem, where the associated mapping consists of two parts: an approximation mapping satisfying a p-th-order Lipschitz bound with respect to the original mapping, and the gradient mapping of a (p+1)-th-order regularization. The optimal global convergence is guaranteed for the ARE method, while a p-th-order superlinear local convergence is shown to hold if the VI is strongly monotone. With the general concept of the approximation mapping, a variety of first-order methods as well as high-order methods can be formulated as specialized instances of the ARE framework. In this talk, we specifically demonstrate a second-order example of the ARE framework for solving monotone VIs. While the subproblem in each iteration involves solving another monotone VI, we show how it can be estimated by first-order subroutines by reducing the subproblem to simpler gradient projection steps. In particular, the technique used in this reduction is exactly an application of the ARE scheme.

3 - Barrier-based second-order method for bilevel optimization

Xiaotian Jiang, University of Minnesota, Minneapolis, MN, United States, Jiaxiang Li, Shuzhong Zhang, Mingyi Hong

In this work, we propose algorithms for solving a class of Bilevel Optimization (BLO) problems, with applications in areas such as signal processing, networking, operations research, and machine learning. Specifically, we develop a novel barrier-based second-order method that transforms the inequality-constrained BLO problem into an unconstrained problem. For the reformulated problem, we compute the implicit gradient and Hessian, and develop a cubic Newton optimization scheme, which involves a single cubic Newton step and the (approximate) solution of the linearly constrained convex lower-level task at each iteration. Under certain mild assumptions, we establish non-asymptotic convergence guarantees of the proposed method to second-order stationary points. Finally, we conduct a number of experiments to demonstrate the potential of the proposed algorithm.

4 - Frank-Wolfe type algorithms for non-convex block optimization models

Casey Garner, University of Minnesota, Minneapolis, MN, United States, Shuzhong Zhang

We present new Frank-Wolfe type algorithms for several non-convex block optimization models with convergence guarantees to first-order stationary and KKT points. Our numerical experiments showcase our method outperforms ADMM when convexity is lost. Additionally, we present new notations of stationarity which inspire novel approaches for locating first-order KKT points for non-convex block optimization models.

TA32

Summit - 423

Multi-stage Optimization under Risk Measures

Invited Session

OPT: Global Optimization

Chair: Renyuan Xu, University of Southern California, Los Angeles, CA, United States

Co-Chair: Luhao Zhang, Columbia University, New York, NY, United States

1 - Causal Distributionally Robust Optimization -- Duality and Sensitivity

Yifan Jiang, University of Oxford, Oxford, United Kingdom

We study causal distributionally robust optimization (DRO) problems in a discrete-time dynamic setting where the model uncertainty is quantified using causal optimal transport. This discrepancy fully characterizes the spatial and temporal information of stochastic processes. Our approach focuses on two fundamental aspects: duality theory and sensitivity analysis. We establish a tractable dynamic duality for general penalized DRO. Additionally, we derive an explicit formula for the distributional sensitivity of value function to model uncertainty. We extend classical results to the so-called adapted objective functions. We illustrate our theoretical results with applications such as distributionally robust version of average value-at-risk, stochastic control, and optimal stopping.

2 - Multi-Period Compliance Mean Field Game Problems

Steven Campbell, Columbia, New York, NY, United States, Yichao Chen, Sebastian Jaimungal

We introduce a mean field game framework for multi-period inventory compliance problems with endogenous price formation. Our model allows firms to bank their inventory in order to meet obligations in subsequent periods and has applications to renewable energy certificate markets where quotas may be imposed by regulators. Under suitable conditions we prove that the representative firm's problem is convex and admits a unique solution for fixed initial data. Then, using variational arguments we characterize the mean field equilibrium in terms of a temporally and spatially coupled system of MV-FBSDEs. We conclude by illustrating numerical solutions to several examples and analyzing the properties that they exhibit.

3 - Beyond Absolute Continuity: a New Class of Dynamic Risk Measures

Luhao Zhang, Columbia University, New York, NY, United States

The modern theory of risk measures copes with uncertainty by considering multiple probability measures. While it is often assumed that a reference probability measure exists, under which all relevant probability measures are absolutely continuous, there are examples where this assumption does not hold, such as certain distributional robust functionals. In this talk, we introduce a novel class of dynamic risk measures that do not rely on this assumption. We will discuss its convexity, coherence, and time consistency properties.

4 - Optimal Sequential Hypothesis Testing with Costly Information Acquisition

Haotian Zong, Johns Hopkins University, Baltimore, MD, United States, Renyuan Xu, Luhao Zhang

We study the sequential testing problem of two alternative hypotheses regarding an unknown parameter when the information is costly. Mathematically, we frame it as a sequential optimal stopping problem with two different informational regimes. This model not only enhances the understanding of the dynamics between cost and information in hypothesis testing but also offers practical insights for adaptive and non-adaptive testing frameworks.

5 - Risk-sensitive Markov Decision Process and Learning under General Utility Functions

Zhengqi Wu, University of Southern California, Los Angeles, CA, United States, Renyuan Xu

Reinforcement Learning (RL) has gained substantial attention across diverse application domains and theoretical investigations. Existing literature on RL theory largely focuses on risk-
neutral settings where the decision-maker learns to maximize the expected cumulative reward. However, in practical scenarios, decision-makers often persist in heterogeneous risk preferences subject to outcome uncertainties, which can not be well-captured by the risk-neural framework. Incorporating these preferences can be approached through utility theory, yet the development of risk-sensitive RL under general utility functions remains an open question for theoretical exploration.

In this talk, we consider a scenario where the decision-maker seeks to optimize a general utility function of the cumulative reward in the framework of a Markov decision process (MDP). To facilitate the Dynamic Programming Principle and Bellman equation, we enlarge the state space with an additional dimension that accounts for the cumulative reward. We propose a discretized approximation scheme to the MDP under enlarged state space, which is tractable and key for algorithmic design. We then propose a modified value iteration algorithm that employs an epsilon-covering over the space of cumulative reward. In the case with access to a simulator, we demonstrate that our algorithm can efficiently identify a near-optimal policy with a sample complexity guarantee. In the case without an accessible simulator, our algorithm utilizes an upper-confidence-bound type of exploration approach, achieving a regret guarantee. For both of the algorithms, we match the theoretical lower bounds for the risk-neutral setting in terms of the total number of time steps.

TA33

Summit - 424

Recent Advance in Optimization Under Uncertainty: Theory and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Man Yiu Tsang, Lehigh University, Bethlehem, PA, United States

Co-Chair: Karmel Shehadeh, Lehigh University, Bethlehem, PA, United States

1 - Distributionally Robust Hydrogen Network Expansion Planning with Decision-Dependent Uncertainty

Beste Basciftci, Tippie College of Business, University of Iowa, Iowa City, IA, United States, Sezen Ece Kayacik, Albert Schrotenboer, Ms I F A Vis, Evrim Ursavas

The transition to a green hydrogen economy faces a critical challenge known as the 'chicken-and-egg dilemma,' wherein establishing a hydrogen supply network relies on demand, while demand only grows with reliable supply. In addition, as the hydrogen market is in the early stage, predicting demand distributions is challenging. This study addresses these complex issues through the introduction of a distributionally robust decision-dependent hydrogen network expansion planning problem. The problem determines optimal locations and production capacities of the suppliers considering the moments of the stochastic hydrogen demand as a function of these investment decisions. We provide two different reformulations that consider continuous and discrete support sets. To efficiently solve the reformulations, we develop a tailored algorithm based on column-and-constraint generation. We enhance the computational performance through solving master problems to a relative optimality gap, decomposing the subproblems and heuristic warm-start cuts. Comparative experiments demonstrate the superiority of our algorithm over the classical column-and-constraint generation algorithm and Benders decomposition. To validate the effectiveness of our approach in decision-making, we create a case study focusing on the Hydrogen Energy Applications in Valley Environments for Northern Netherlands. In turn, we provide insights to policymakers for strategic development of hydrogen network, overcoming the 'chicken-and-egg dilemma' and uncertain hydrogen market conditions.

2 - Data-Driven Multistage Stochastic Optimization on TIME Series

Rohit Kannan, Virginia Tech, Blacksburg, VA, United States, Nam Ho-Nguyen, Jim Luedtke

We study a data-driven framework for multistage stochastic optimization assuming only access to a single historical trajectory of the underlying stochastic process. Conditional on the most recent observations of the stochastic process, the goal is to determine a decision policy that minimizes the expected cost over the next T time periods. We investigate a data-driven approximation in which a time-series model is fit based on the historical data, and then the residuals from this fit are used to build a discrete approximation of the stochastic process. Our formulations are flexible and can accommodate nonlinear VARX and multivariate GARCH models, and the approximation we study can be solved via the stochastic dual dynamic programming algorithm. We derive conditions on the underlying stochastic process, the time series model, and the optimization model under which solutions to the approximation possess asymptotic and finite sample guarantees. Experiments on a hydrothermal scheduling model illustrate the potential benefits of our data-driven formulations even when the time series model is misspecified.

3 - On the Trade-off Between Distributional Belief and Ambiguity: Conservatism, Bias, and Asymptotic Properties

Man Yiu Tsang, Lehigh University, Bethlehem, PA, United States, Karmel Shehadeh

We present a new data-driven trade-off (TRO) approach for modeling uncertainty that serves as a middle ground between the optimistic approach, which adopts a distributional belief, and the pessimistic distributionally robust optimization approach, which hedges against distributional ambiguity. We equip the TRO model with a TRO ambiguity set characterized by a size parameter controlling the level of optimism and a shape parameter representing distributional ambiguity. We first show that constructing the TRO ambiguity set using a general star-shaped shape parameter with the empirical distribution as its star center is necessary and sufficient to guarantee the hierarchical structure of the sequence of TRO ambiguity sets. Then, we analyze the properties of the TRO model, including quantifying conservatism, quantifying bias, and establishing asymptotic properties. Specifically, we show that the TRO model could generate a spectrum of decisions, ranging from optimistic to conservative decisions. Additionally, we show that it could produce an unbiased estimator of the true optimal value. Furthermore, we establish the almost-sure convergence of the optimal value and the set of optimal solutions of the TRO model to their true counterparts. We exemplify our theoretical results using stylized optimization problems.

4 - Distributionally Robust Optimization with Multimodal Decision-Dependent Ambiguity Sets

Xian Yu, The Ohio State University, Columbus, OH, United States, Beste Basciftci

We consider a two-stage distributionally robust optimization (DRO) model with multimodal uncertainty, where both the mode probabilities and uncertainty distributions could be affected by the first-stage decisions. To address this setting, we propose a generic framework by introducing a φ-divergence based ambiguity set to characterize the decision-dependent mode probabilities and further consider both moment-based and Wasserstein distance-based ambiguity sets to characterize the uncertainty distribution under each mode. We identify two special φ-divergence examples (variation distance and chi-square-distance) and provide specific forms of decision dependence relationships under which we can derive tractable reformulations. Furthermore, we investigate the benefits of considering multimodality in a DRO model compared to a single-modal counterpart through an analytical analysis.
We provide a computational study over the facility location problem to illustrate our results, which demonstrate that omission of multimodality and decision-dependent uncertainties within DRO frameworks result in inadequately performing solutions with worse in-sample and out-of-sample performances under various settings.

TA34

Summit - 425

Data-driven Approaches in Reliability Analysis

Contributed Session

Chair: Ziyu Wang, N/A

1 - Feature Extraction to Enhance Fault Detection in Electro-Mechanical Systems Accounting for Component Degradation

Hyunsik Yoon, Pohang University of Science and Technology, Pohang, Korea, Republic of, Duck Young Kim

Many preprocessing techniques for fault detection in electro-mechanical systems have been studied to pre-analyze different types of data collected from various sensors. Among these, feature extraction techniques from large and complex data are essential to reduce dimensionality and improve model performance. For example, Principal Component Analysis (PCA), Multi-Dimensional Scaling (MDS), Linear Discriminant Analysis (LDA), and Neural-Network (NN) based feature extraction methods have been effectively used to provide basis information about certain data characteristics such as covariance, data frequency, entropy, and labels. This paper presents a novel feature extraction method for fault detection in electro-mechanical systems in cases where degradation information is available. To do this, a multi-modal classification model integrating time series and discrete sensor data is developed to discern system degradation levels. The effectiveness of the proposed feature extraction method for fault detection is demonstrated through two case studies: an injection molding system and a bearing-shaft system.

2 - Optimal control of failing systems under deteriorating measurements

Bin Liu, University of Strathclyde, Glasgow, United Kingdom, Guo Shi, Lesley Walls

Condition-based maintenance (CBM) employs sensors and communication systems to monitor equipment conditions and present system failures based on the measurements. In CBM literature, it is often assumed that sensor readings are either perfect or only subject to measurement errors. However, this overlooks the fact that sensors, being electromechanical systems, are also prone to degradation over time due to environmental and aging factors. This degradation can compromise the accuracy of the measurements and pose challenges for maintenance scheduling. To address these challenges, we propose a three-layer factorial hidden Markov model (HMM) designed to understand the dynamics of hidden states—specifically sensor degradation, system degradation, and the translation from these hidden states to observable data. Unlike traditional HMMs, which typically utilize a complex $K^2 \times K^2$ transition matrix (K is the number of states), our model simplifies this structure into two distinct K×K matrices, significantly reducing computational demands. Additionally, we have developed an enhanced expectation–maximization algorithm to iteratively refine the model parameters based on historical monitoring data. Subsequently, we formulate the maintenance problem into a partially observed Markov Decision Process over an indefinite horizon. This optimal maintenance strategy is characterized by a control limit policy that is defined by a specific belief threshold. The effectiveness of the proposed approach is validated through numerical examples and a practical case study. This model not only improves maintenance efficiency but also enhances the reliability of the monitoring system by accounting for sensor health in the maintenance scheduling process.

3 - Condition-Based Maintenance for Degradation Systems with Quality Characteristics

Ziyu Wang, Tianjin University, Tianjin, China, People's Republic of, Xiujie Zhao

With the development of surveillance technology, especially the improvement of sensors, condition-based maintenance (CBM) has gained an increasingly popularity. However, the growing flexibility of manufacturing systems poses challenges for the accurate assessment of system states, which makes it difficult for traditional CBM to leverage its advantages. The paper investigates the optimal maintenance strategy for manufacturing systems with the consideration of state assessment uncertainties. The system consists of the production part and the output part, such as machines and products. The underlying states of machines cannot be observed directly, but can be revealed through the information of product quality. Inspections are carried out in batches and the results are used for the assessment of the underlying state of machines. The optimization model is formulated as Partially Observable Markov Process (POMDP) with the optimal maintenance policy obtained by minimizing the expected maintenance cost in both finite and infinite horizons. Point Based Value Iteration (PBVI) algorithm is used to solve the POMDP problem and to compute the expected maintenance cost. Finally, we give numerical examples to illustrate the effectiveness of the proposed model and conduct sensitivity analyses to explore the impact of parameters on the model.

TA35

Summit - 427

Data Mining Society (DMS)

Invited Session

Data Mining

Chair: Ke Yang, UNIVERSITY OF ARKANSAS AT FAYETTEVILLE, Fayetteville, 72701, United States

1 - Deep Learning Applied to Facility Inspection and Health Monitoring Using Images Captured by Uav

Ke Yang, University of Arkansas, Fayetteville, AR, United States

Maintaining sustainable, livable cities for economic growth is one of the biggest urban challenges. As part of the related efforts, it is desired to quickly detect and fix leaks in roofs that cover a large portion of building areas. However, due to the limitations of traditional inspection technologies, roof leakage caused by materials deterioration is hard to detect until leaks with various shapes, sizes, and propagation directions become visible. To overcome this practical challenge, we proposed a novel MAE-YOLOv8 model to recognize percolation features from images data captured by Unmanned Aerial Vehicles (UAVs). Specifically, the attention mechanism is embedded in a deep neural network to focus on subtle features and enhance the extraction performance in the backbone. Meanwhile, more connections are established in the neural network architecture by convolution and concatenation operations to acquire comprehensive information from the input data. Finally, the fused features are fed into the detectors for subject detection. The proposed model enables cost-effective roof leakage detection using UAV imagery. A case study including both extensive and ablation experiments demonstrates the model's performance with respect to detection accuracy and robustness. For practitioners, the model not only provides a proactive approach to urban resiliency but also promotes applications of computer vision in broader areas.

2 - A Mixture of Decomposition-based Transformer Experts Network

Yu Jin, University at Buffalo, Buffalo, NY, United States

Time series forecasting plays a crucial role in decision-making, control, and monitoring across various fields. However, accurately predicting nonstationary time series remains a challenge because traditional methods struggle to capture the complex temporal dynamics. Recent research into Transformer models for time series forecasting has showcased their ability to manage long-term dependencies. In this study, a Mixture of Decomposition Transformers Experts Network (MixDecformer) is introduced to combine the Transformer architecture with a decomposition method to capture varying levels of detail in the decomposition components and mitigate the "end effect" problem often associated with decomposition methods.

3 - Predictive Study on the Cutting Energy Efficiency of Dredgers Based on Cutting Specific Energy

JUNLANG YUAN, Wuhan University of Technology, Wuhan, China, People's Republic of, Shidong Fan, Ke Yang

Abstract: The cutter suction dredger (CSD) serves as a crucial tool in port and waterway construction. The cutter, an essential component of the CSD, directly impacts the construction progress, safety, and economic benefits. The cutting specific energy of the cutter represents the energy expended to cut rock per unit volume and is closely related to the cutter's rock breaking efficiency. This study establishes a computational model for the cutting specific energy of the cutter; based on actual engineering construction data, it introduces a method for predicting the cutting specific energy of the cutter using neural network algorithms and verifies the method through engineering examples.

TA36

Summit - 428

Sustainable Operations in Urban Logistics

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Lei Zhao, Tsinghua University, Beijing, N/A

Co-Chair: Shu Zhang, Chongqing University, Chongqing

1 - Assessing the Effect of Consolidation of Freight Transport: A Case Study in Norway

Julio Goez, NHH Norwegian School of Economics, Bergen, Norway, Cosku Can Orhan, Mario Guajardo, Stein Wallace

The growing urban population and increasing e-commerce sales cause more and more parcel deliveries within cities. Motivated by these trends, in this project we cooperate with public authorities of a city in Norway to study the effect of consolidation of freight transportation. The effect is measured along metrics on distance and number of stops. To compute these metrics, we deploy a three-stage methodology, which involves clustering, capacitated vehicle routing, and split delivery problems. We use data on road network and population from the public authority, and on deliveries from a major carrier. We present results for a number of data instances, varying the number of companies, their size and the degree of consolidation. Ultimately, our work aims at providing city authorities with a tool for the assessment of freight transport regulations.

2 - From Resisting Algorithm to Feeding Algorithm: Strategic Behavior of Couriers

Chen Wang, Tsinghua University, Beijing, China, People's Republic of, Junjie Xiong, Lei Zhao, Jan Fransoo

This study investigates the mechanisms behind couriers' resistance and feeding behaviors under the algorithmic control of online platforms. Utilizing grounded theory and focused semi-structured interviews, we detail how couriers interpret, resist, and occasionally manipulate algorithmic controls. Our analysis aims to reveal the specific mechanisms that give rise to the ubiquitous resistance behaviors. We demonstrate how resistance transitions into "feeding" the algorithm with intentional responses and data for favorable job assignments and compensation. Based on these understandings, we propose integrating a well-being payoff into the algorithm objective, which shifts from enforcing compliance to aligning expectations and enhancing couriers' person-environment fit.

3 - Urban Delivery with Hierarchical Warehouses

Tirui Cao, Tsinghua University, Beijing, China, People's Republic of, Xue Luo, Yilei Wan, Hengle Qin, Chen Wang, Lei Zhao

To fulfill orders with various delivery urgency levels, a business-to-customer (B2C) e-commerce company employs a hierarchical warehouse logistics system, with a city distribution center (CDC) located in the suburban area and several (much) smaller forward warehouses located in the city area. The CDC is further from the last-mile stations (and requires longer delivery time) but normally has adequate inventory, whereas the forward warehouses are closer to the last-mile stations (and require shorter delivery time) but are restrained to limited inventory. In this paper, we study the daily delivery routing problem of this logistics system's first-tier transportation (i.e., the CDC or forward warehouses to the last-mile stations). We formulate this problem as a binary integer program and develop a branch-price-and-cut (BPC) algorithm, which consists of an efficient label-setting pricing algorithm and a tailored branching scheme. A comprehensive numerical study demonstrates the effectiveness of the proposed BPC algorithm and provides insights into the value of utilizing hierarchical warehouses in urban delivery.

4 - Impact of Congestion Pricing Schemes on Depot Location for Last-mile Delivery

Rui Liu, School of Economics and Business Administration, Chongqing University, Chongqing, China, People's Republic of, Shu Zhang, Ann Campbell

As urbanization increases, municipalities are concerned about traffic congestion and air pollution caused by road-based transportation. To reduce congestion and pollution in city centers, several cities have introduced congestion charging schemes to discourage vehicles from entering the inner city. The list includes London, Stockholm, Milan, Gothenburg, Valletta (Malta), Durham, Singapore, and New York City (to be started on June 30, 2024). The charging types include per-day, per-entry, per-gantry, and per-minute fees. As far as we are aware, little research has been done to examine the impact of congestion charges on logistics operations. In this study, we investigate how different charging schemes will impact the depot location in a last-mile delivery problem. We consider different geographies and identify conditions under which optimal routes with the candidate depot can be obtained in polynomial time. We find out that the congestion pricing schemes can shift the choice of depot location. That is, with different congestion charge types and rates, the optimal depot location may vary from inside the congestion zone to outside the zone, or from near the customers to far away from the customers. We propose a heuristic algorithm to find good quality depot locations for large-size problems considering different congestion pricing schemes.

5 - Learning for multiple purposes: A Q-learning enhanced hybrid metaheuristic for parallel drone scheduling traveling salesman problem

Ping Chen, Nankai University, Tianjin, China, People's Republic of, Qianlong Wang

In recent years, the integration of artificial intelligence (AI) techniques with traditional operations research (OR) methods has emerged as a significant trend for addressing combinatorial optimization problems. This paper presents a novel hybrid heuristic approach, Q-learning aided Slack Induction by String Removals (QSISRs), which combines reinforcement learning with a ruin-and-recreate metaheuristic to enhance algorithmic performance. The focus of our study is on the Parallel Drone Scheduling Traveling Salesman Problem (PDSTSP), a contemporary challenge in urban logistics distribution where drones are increasingly being utilized alongside traditional truck and courier services for last-mile delivery. The PDSTSP is characterized by a scenario where a portion of customers (ranging from 20% to 100%) are within the flight range of drones, enabling direct service from a central depot without the need for synchronization with truck operations. To evaluate the efficacy of the QSISRs, we applied it to the PDSTSP and conducted a series of numerical experiments. The results not only validate the positive impact of incorporating Q-learning into the metaheuristic framework but also demonstrate the robustness of QSISRs in effectively solving the PDSTSP. Furthermore, comparative analysis with existing heuristic approaches underscores the superiority of QSISRs, positioning it as one of the state-of-the-art methods for tackling this complex problem in the field of urban logistics.

TA37

Summit - 429

Advancements in Marine Shipping Operations

Contributed Session

Chair: Jianxia Liu, Harbin Engineering University

1 - Decision Support for Zero-Impact Marine Shipping

Simone Philpot, University of British Columbia, Vancouver, BC, Canada, Zixuan Liu, Hugo Ricart, Clara Kaufmann, Terre Satterfield, Amanda Giang

Marine shipping enables trade, mobility, food security, emergency services, and national security. While marine vessels are the most efficient commercial transport option available in terms of CO2 emissions per cargo tonne/mile, their environmental impacts are significant. Viable and sustainable marine shipping, from both social and ecological perspectives, is an essential sustainability goal. Many technologies for reducing vessels' environmental impacts are available and in development, each with distinct trade-offs, risks, and uncertainties. Collaborating with Clear Seas, a non-profit organization, and National Research Council Canada (NRCC), we are developing decision support for evaluating vessel technologies that capture trade-offs and co-benefits across a suite of socio-environmental impacts. Here, we present findings from case studies with the Canadian Coast Guard and the NRCC. Combining expert interviews from government, industry, and consulting, with structured decision-making workshops, we identify key constraints, objectives and performance indicators around design, procurement, and research and development. For example, we find that opportunities to innovate are unevenly distributed between commercial vessels and those providing critical services, such as ferries and ice breakers. We also document challenges associated with alternative technologies around crew safety and possible new training requirements. These types of considerations impact technology feasibility and the incentives needed to support innovation. We integrate these types of elicited perspectives into an evaluative framework capturing co-benefits and trade-offs of technologies as their impacts cascade through social and ecological systems. This work contributes to a growing role for decision support as the marine shipping sector navigates changing regulatory demands and evolving technological opportunities.

2 - A Hybrid ARIMA-LSTM Approach for Improved Forecasting of Marine Bunker Prices

Orestis Schinas, University of Aegean, Athens, Greece, Theano Krassa

In the face of the maritime industry's transition towards sustainability, precise forecasting of bunker fuel prices is crucial. This study addresses a significant gap in existing forecasting methodologies for maritime fuels, a sector that has lagged in adopting advanced predictive analytics. Our research is motivated by the urgent need for more accurate forecasting methods, pivotal for managing fuel costs amid global efforts to green the shipping sector.

Based on a relevant research paper, building on the 2015 WMU study by Stefanakos and Schinas (Stefanakos & Schinas, 2015), which applied fuzzy time series to bunker price forecasting, this paper introduces a novel hybrid model that combines AutoRegressive Integrated Moving Average (ARIMA) with Long Short-Term Memory (LSTM) networks. Our model significantly improves the prediction accuracy for various types of bunker fuels—High Sulfur Fuel Oil (HSFO), Very Low Sulfur Fuel Oil (VLSFO), Marine Diesel Oil (MDO), and Marine Gas Oil (MGO)—in major ports such as Rotterdam, Fujairah, Singapore, and Houston. Importantly, this study pioneers the integration of fuzzy logic and LSTM in forecasting prices for Liquefied Natural Gas (LNG).

The methodology focuses on the statistical properties of time series, enhancing forecasting precision and enabling better strategic planning for shipping companies. This research not only advances academic knowledge but also equips the maritime industry with practical tools for navigating the sustainability transition.

3 - Time-Band Network Based Vessel Schedule Recovery in Liner Shipping Services Under Sulfur Emission Regulation

Jianxia Liu, Harbin Engineering University, Harbin, China, People's Republic of, Yuzhen Hu

To reduce sulfur emissions from ships, the International Maritime Organization (IMO) has established sulfur emission control areas (SECA), which affect liner ships in operation. Improving resilience after disruptions in operation while complying with SECA regulations is challenging for liner shipping. Therefore, this paper proposes a multi-arc reduced time-band network model for the operational vessel schedule recovery problem in liner shipping services, jointly minimizing navigation and disruption recovery costs following sulfur emission regulations. Besides, a marking method and two arc-reduced strategies are adopted in this study to improve the performance of the time-band network model, with corresponding theorem analysis. Computational experiments based on an actual liner shipping route demonstrate the effectiveness of the proposed network model, marking method, and arc-reduced strategies. This study's experimental results and management insights can provide decision-making references for the liner shipping industry, enhancing maritime transportation resilience under environmental requirements.

TA38

Summit - 430

Resilience in Aviation

Contributed Session

Aviation Applications

Chair: Felipe Delgado, N/A

1 - Navigating Turbulence: Optimizing Crew and Financial Resilience by Considering Disruptive Events

Burak Cankaya, Embry Riddle Aeronautical University, Daytona Beach, FL, FL, United States, Burak Cankaya, Eyyub Kibis, Bulent Erenay, Aaron Glassman, Ozlem Cosgun

The airline industry, known for its susceptibility to volatility, faces the continuous challenge of balancing long-term planning with the agility to respond to disruptive events. These disruptions demand a reassessment of network and fleet strategies, pilot hiring, and career progression plans, aiming for a harmony that benefits the airline and its pilots. This presentation delves into a mathematical model designed to optimize pilot resource allocation during disrupted routine operations. We'll unveil a systematic approach to pilot manpower planning, ensuring airlines are prepared to reallocate pilots among unforeseen changes efficiently. The talk will further explore a financial simulation that scrutinizes four distinct pilot allocation strategies, salary adjustments, and lay-off scenarios, assessing their viability and impact on the airline's financial health. Concluding with actionable insights, the study presents valuable case studies and recommendations for industry stakeholders, emphasizing strategies for reducing uncertainties in pilot planning. Moreover, it suggests avenues for future research to strengthen the airline's financial recovery post-disruption, ultimately fostering a more resilient airline industry.

2 - Improved Airline Network Topologies for Robust Aviation Infrastructures

Ashwini Ravindran, Texas A&M University, College Station, TX, United States, Sergiy Butenko

We propose new heuristic models for an optimal airline network design which can provide cost-efficient strategies for boosting operational efficiency and enhanced network connectivity, while also strengthening resilience against disruptions at key nodes. These newly developed theoretically optimal network designs are then compared with the existing airline networks to evaluate its performance with regard to robustness and tolerance to targeted attacks.

3 - Aircraft Maintenance Scheduling Under Uncertain Task Processing TIME

Felipe Delgado, Pontificia Universidad Católica de Chile, Santiago, Chile, Matías Villafranca, Mathias Klapp

Delays in aircraft maintenance tasks can lead to costly disruptions in airline operations, causing flight delays and significantly impacting operational efficiency and expenses. This study addresses the challenge of uncertainty in maintenance task processing times and designs cost-efficient aircraft maintenance schedules via two-stage stochastic programming. In the first stage, we determine the daily tasks to execute in each aircraft, specifying their start time and assigned maintenance base and technician. In the second stage, adjustments are made to task start times and flight departure times based on realized task processing times. The objective is to minimize costs associated with outsourced maintenance tasks, technician overtime, and flight delays. To tackle this problem, an ad-hoc Adaptive Iterated Local Search heuristic is developed, which efficiently evaluates second-stage costs while exploring first-stage solutions. The proposed approach is tested using computationally simulated instances. Results indicate significant cost savings compared to deterministic and conservative approaches, with average savings of 77% and 45%, respectively. Additionally, the proposed model outperforms a benchmark solution that optimizes time buffers by achieving 12% average cost savings. Furthermore, the study investigates the cost implications of varying structural parameters such as task granularity, processing time variability, workload, and cost structure.

TA39

Summit - 431

Railway Freight Operations and Network Management Optimization and Modelling

Invited Session

Railway Applications

Chair: Tyler Dick, University of Texas at Austin, Austin, TX, United States

1 - Causal Inference Study on the Impact of Distributed Power on Derailment Severity

Xinhao Liu, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Chen-Yu Lin, Christopher Barkan

Train derailment characteristics have been extensively analyzed. In existing studies, few of them pay attention to distributed power (DP) trains. DP trains, which can handle heavier tonnages and operate on challenging routes, have become increasingly prevalent in the U.S. over the past 15 years. This research contrasts the derailment risks associated with DP and conventional (non-distributed power) freight trains, aiming to enhance industry understanding of DP train operations.

This paper presents an exploratory and statistical analysis on the effects of DP on freight train derailments, focusing on Class I mainline and siding derailments from 2014 to 2023 across North America. We examine derailments of unit and manifest trains with and without DP, analyzing derailment characteristics, frequencies, and severities. Regression models quantify relationships, and the most frequent and severe causes of derailments for each train group are discussed.

Preliminary results indicate that DP trains, predominantly unit trains, exhibit higher derailment speeds, longer train lengths, and heavier weights compared to their non-DP counterparts, although the difference in the percentage of vehicles derailed is not statistically significant. This analysis provides insights into the safety profiles of DP trains relative to conventional freight trains.

Additionally, I aim to conduct a causal inference analysis on the impact of distributed power on derailment severity using regression methods and propensity score matching. While results are pending, this forthcoming analysis will further elucidate the causal relationships and support more advanced risk management strategies for freight train operations.

2 - Route and Schedule Optimization Method of the Mixed Train with “Empty-Loaded-Train” Service Network

Zhicong Yuan, Beijing Jiaotong University, Beijing, China, People's Republic of, Haiying Li, Ying Wang, Jiarui Zhang, Chao Yu

This study centers on a specific type of mixed train, which enables loaded wagons and empty wagons to be grouped together. To effectively organize loaded wagons and distribute empty wagons, we propose a method for optimizing the route and schedule of the mixed train. Firstly, the qualitative analysis of the train operation process in China Railway Special Cargo Logistics Co., Ltd (CRSCL) is conducted. And we propose the feasibility and potential benefit of the train. Secondly, we design a three-layer "Empty-Loaded-Train" time-space network to describe the empty wagon distribution, the loaded wagon transportation, and the train organization. Then, a multi-commodity flow model is proposed to optimize the route and schedule of the train for realizing the potential benefit. Furthermore, a sensitivity analysis is conducted in the scene of CRSCL. The results demonstrate that the delivery time of loaded wagons is shortened by 39.56% and an additional 44.44% of empty wagons distribution task is completed.

3 - Aggregated Fundamental Diagrams for Rail Traffic Flow

Han Zheng, Arizona State University, Tempe, AZ, United States, Francesco Corman, Tyler Dick, Xuesong Zhou

This presentation presents a new approach to better characterize rail traffic stream properties by connecting the concepts of Aggregated Fundamental Diagram for considering Corridor-level spatial Heterogeneity (AFD-cv) and Traffic Stream Model for homogeneous rail segments (TSM-h). Its aim is to extend the application of traffic stream models to rail traffic flows, thus bridging the gap between existing traffic flow theories and emerging rail-oriented multi-resolution modeling. Key rail-specific elements, such as block signal controls, nominal maximum speed, and intermediate station stopping plans, are considered in the proposed simplified models, by a few, crucial parameters. The TSM-h model links segment-level throughput with important variables and can recognize different rail traffic state regimes by critical points with structural mathematical equations. Based on TSM-h, the AFD-cv model introduces a train density variation index, which can reflect the effects of service stopping plans or timetables on throughput of the corridor-level system. Subsequently, a microscopic train-following model is derived from the TSM-h model, by leveraging its differentiable and continuous properties. Numerical experiments demonstrated the models' ability to represent traffic flow dynamics consistently across microscopic-to-aggregated and segment-to-corridor resolutions. The proposed models represent a step towards a better characterization of rail traffic flows and offer significant potential for planning and operational improvements.

4 - Empowering Private Rail Freight Operators: Leveraging Data-Driven Predictive Analytics

GAURAV KUMAR, University of British Columbia, Vancouver, BC, Canada, Vancouver, BC, Canada

Rail freight plays a pivotal role in fostering economic growth through its efficient services within supply chains. However, challenges such as single-track sections, manual dispatching, operational fluctuations, accidents, and diverse train characteristics often lead to delays, emphasizing the critical need for accurate Estimated Time of Arrival (ETA) predictions. To address this challenge, our study employs machine learning techniques to develop a data-driven predictive model tailored for private rail freight operators. Specifically, we utilize ensemble modeling, combining multiple linear regression and K-Nearest Neighbor algorithms, trained and validated using data collected from the Indian rail freight operator RFOY between August 2022 and October 2023. Continuously updating our estimates along the journey, we refine predictions of final destination arrival times by gathering data at specific points. Our regression model demonstrates satisfactory performance, exhibiting a strong correlation between predicted and actual arrival times, with over 95% accuracy for the majority of trips. Furthermore, we have optimized RFOY's daily operations by developing an open-source Graphical User Interface dashboard, "RFOY ETA PREDICTOR," which aligns with operational needs, ensuring efficiency throughout the rail system. Managerial recommendations encompass aligning freight train services with revenue goals, optimizing turnaround times, and boosting ETA predictions to enhance operator efficiency and customer satisfaction. These measures aim to secure a competitive edge and readiness for future demand in real-world scenarios.

TA40

Summit - 432

Behavioral Decision-Making in Retail Operations

Invited Session

Behavioral Operations Management

Chair: Han Oh, Tilburg University, Tilburg, Netherlands

1 - Newsvendor Double-Counting Bias in Automatic Grocery Retail Replenishment

Bengü Nur Özdemir, IE Business School, Madrid, Spain, Antti Tenhiälä

This paper investigates why store managers of a supermarket chain order more than their automatic store replenishment (ASR) system's order proposals in inventory replenishment decisions. We empirically show that store managers are susceptible to a novel bias when increasing ASR orders before finalizing the replenishment decisions: newsvendor double-counting bias. This bias suggests that decision makers may apply the newsvendor logic mentally when they encounter an algorithmic proposal as if this proposal refers to a demand forecast. Yet, in the context of the usage of ASR systems, this would be a double-counting bias because the newsvendor optimization logic is applied to forecasts by the algorithm before those forecasts are converted into order proposals. We analyze ordering decisions of store managers of our collaborator supermarket chain by a probit model with sample selection model, a Heckman selection model, and an endogenous treatment regression model. All these models include exclusion restrictions adopted from the literature to tackle the endogeneity. We also control for anchoring and supply line underweighting, and take them as benchmark biases to understand the relative significance of newsvendor double-counting bias. These analyses show that order increases triggered by newsvendor double-counting bias do not lead to additional sales. This shows that the discretionary power of the users to increase orders may lead to increased inventory holding costs and waste.

2 - Behavioral multi-lever decision-making: A study of consumer return policy, price, and inventory decisions

Han Oh, Tilburg University, Tilburg, Netherlands, Huseyn Abdulla, Rogelio Oliva

We investigate consumer return policies recognized and studied by operations management scholars as an important managerial decision in a retail environment. Our research investigates, through randomized experiments, the behavioral aspects of return policy decisions and their interaction with other common operational decisions.

3 - How to Account for Behavioral Newsvendors: The Robust Buyback Contract to Address Response Uncertainty

Christina Imdahl, Eindhoven University of Technology, Eindhoven, Netherlands, Michael Becker-Peth, Kai Hoberg

The normative (expected profit-maximizing) theory assumes that decision-makers are fully rational, but in reality, they deviate from the optimal response. In contract negotiations, contract parameters are often optimized based on the assumption of rational behavior. Deviations from rational behavior can lead to significant costs for all parties involved. To address this, we propose robust optimization to obtain contract parameters that are robust to deviations from rational behavior.
We apply this optimization approach to obtain the robustly optimal contract parameters for the buyback contract when there is limited knowledge of the buyers' responses.
We compare the robust contract to the normative contract if people behave according to known behavioral models, such as mean-anchoring, bounded rationality, or a behavioral model explicitly for the buyback contract. Additionally, we test different robust contracts in a lab experiment with actual human decision-makers. The results demonstrate that a robust contract, based on simple assumptions, can outperform the normative contract in terms of risk while not sacrificing average supply chain profit. This approach offers a practical way to set contract parameters without relying on assumptions about the distribution of responses, except for the support.

4 - Late Payments, Higher Prices? An Experimental Investigation of Competitive Procurement

Kyle Hyndman, The University of Texas at Dallas, Richardson, TX, United States, Matthew Walker

The decision to pay one's suppliers late is commonplace across global supply chains. Motivated by our direct interactions with buyer-side and supplier-side organizations and our analysis of regulatory data, we construct a game-theoretic model and investigate the strategic element to late payments using laboratory experiments. We find that the anticipation of late payments leads to higher supplier prices and lower expected buyer profits. We examine whether an economic incentive (penalty for late payment) deters buyers from reneging upon an agreed payment term. The effectiveness of introducing a penalty for late payment is not obvious, because some buyers may choose to circumvent the penalty by offering a longer payment term up front, and the penalty may influence suppliers' bidding strategies. We find that setting the highest penalty for late payment is an effective deterrent mechanism and leads to the highest expected buyer profit through its effect on supplier competition. Thus, establishing a credible commitment to pay one's suppliers on-time is a cost-efficient managerial strategy for buyers. However, a welfare loss arises if the penalty is not set high enough because it fails to align incentives for a subset of buyers.

TA41

Summit - 433

Stochastic Models and Optimization in Service Systems: Matching, Pricing, Scaling, And Planning

Invited Session

Applied Probability Society

Chair: Sophie Yu, Stanford University, Durham, NC, United States

1 - Greedy Matching of Impatient Agents: The Role of Inventory

Angela Kohlenberg, Northwestern University, Evanston, IL, United States

We study dynamic matching in two-sided markets with heterogeneous and impatient demand and supply, and demand-dependent match rewards. These markets face a trade-off between delaying matches to allow for better options and performing matches quickly to prevent agent abandonment. A greedy matching policy attempts to match agents immediately upon their arrival. A policy that greedily matches agents over the entire market ignores the potential benefits of reserving inventory (waiting agents) for better future matches. However, greedily matching supply over a subset of the highest-reward demand may sacrifice the opportunity to match supply before it abandons.
Our results reveal that the optimality of greedy policies depends on the fundamental questions: can and should excess inventory be held on the supply-side of the market? We derive precise expressions of the market parameters, which are translated into intuitive market characteristics, to answer these questions. A policy is scale optimal if its reward loss due to abandonment is on the same order as the optimal policy. We develop an algorithm for general two-sided networks that identifies whether a greedy policy is scale optimal, and, if so, specifies the optimal market design such that greedily matching agents over specified partitions of the network is scale optimal.

2 - Distributed Tate Scaling in Large-Scale Service Systems

Debankur Mukherjee, Georgia Tech, ISyE, Atlanta, GA, United States, Daan Rutten, Martin Zubeldia

In this talk, we will consider a large-scale parallel-server system, where each server independently adjusts its processing speed in a decentralized manner. The objective is to minimize the overall cost, which comprises the average cost of maintaining the servers' processing speeds and a non-decreasing function of the tasks' sojourn times. The problem is compounded by the lack of knowledge of the task arrival rate and the absence of a centralized control or communication among the servers. We draw on ideas from stochastic approximation and present a novel rate scaling algorithm that ensures convergence of all server processing speeds to the globally asymptotically optimum value as the system size increases. Apart from the algorithm design, a key contribution of our approach lies in demonstrating how concepts from the stochastic approximation literature can be leveraged to effectively tackle learning problems in large-scale, distributed systems. En route, we will also discuss the performance of a fully heterogeneous parallel-server system, where each server has a distinct processing speed, which might be of independent interest.

3 - Electric Vehicle Fleet and Charging Infrastructure Planning

Sushil Mahavir Varma, Georgia Tech, Atlanta, GA, United States, Francisco Castro, Siva Theja Maguluri

We analyze an optimal electric vehicle (EV) fleet and charging infrastructure capacity planning problem in a spatial setting. As customer requests arrive at rate λ, the system operator must determine the minimum number of vehicles and chargers for a given service level along with a matching and charging policy that maximizes the service level. We provide a sharp characterization of the fleet size and the charging infrastructure requirements as the demand grows. While a system in which charging times are negligible needs extra $\Theta(\lambda 2/3)$ vehicles on top of the nominal capacity, we show that an EV system has a fundamentally different scaling. Due to charging times, the nominal capacity of the system is increased, but this increased capacity allows for an optimal EV dispatching policy to result in an extra fleet requirement of only $\Theta(\lambda\nu)$ for $\nu \in (1/2, 2/3]$, depending on the number of charging stations and the size of the EV battery packs. We propose the Power-of-d dispatching policy, which achieves this performance by selecting the d closest vehicles to a trip request and choosing the one with the highest battery level. We conduct detailed simulations that verify our scaling results. Lastly, we discuss how to extend our results to accommodate time-varying cyclic demand patterns.

TA42

Summit - 434

Recent Advances in Approximate Dynamic Programming

Invited Session

Applied Probability Society

Chair: Parshan Pakiman, Chicago Booth, 5751 S Woodlawn Ave, Chicago, IL, 60637, United States

Co-Chair: Daniel Adelman, University of Chicago, Booth School of Business, Chicago, IL, United States

1 - Unit Commitment Without Commitment: Managing An Integrated Energy System Under Uncertainty

David Brown, Duke University, Durham, NC, United States, James Smith

Renewable energy sources like solar and wind are prone to variability and uncertainty. Although variable and uncertain demand has always been an issue for power systems, the growing reliance on renewable energy increases the need for system operators to manage production systems carefully. Current industry practice at these utilities involves solving ``unit commitment'' and ``economic dispatch'' optimization problems to choose production plans: these models, while complex, do not explicitly incorporate uncertainty. In this work, we develop an analytical framework based on weakly coupled stochastic dynamic programming to help system operators manage production under uncertainty. We develop relevant theory and demonstrate this dynamic framework using data from the Duke Energy Carolinas and Progress systems. Our numerical experiments demonstrate that this dynamic approach is computationally feasible at an industrial scale. Compared to current practice, our approach is computationally efficient and substantially improves operational efficiency, with the efficiency gains increasing in the renewable capacity of the system.

2 - Multi Shot Linear Programming Approximation of Average Cost Markov Decision Processes

Selva Nadarajah, University of Illinois Chicago, College of Business Administration, Chicago, IL, United States, Parshan Pakiman

Approximate linear programming is a foundational approach to obtain control policies for large-scale average cost Markov decision processes (MDPs) arising in domains such as queuing control, inventory management, and healthcare operations. Its most common use involves approximating the average cost MDP by a perturbed MDP connected to a discounted cost model and then applying approximate linear programming to the resulting perturbed MDP. In contrast, a pioneering approximate linear programming approach directly approximates the average cost MDP in two phases but has been largely overlooked because it cannot guarantee good policy performance. We revisit this pioneering approach and address its limitations by combining random features and multiple approximation attempts. We establish a convergence rate for phase one lower bound to reach the optimal policy cost with a high probability as more random features are sampled. Given a near-optimal phase one lower bound, phase two takes multiple shots at approximating the MDP bias function using a sequence of randomized linear programs that provide policies with improving worst-case performance. Linear programs in this sequence have increasing numbers of random features and are warm-started using information from the immediately preceding linear program. Numerical experiments on joint replenishment and perishable inventory control applications highlight the efficacy of the proposed approach.

3 - Measurized Discounted Markov Decision Processes

Daniel Adelman, University of Chicago, Booth School of Business, Chicago, IL, United States, Alba Olivares Nadal

We build a framework that facilitates the analysis of discounted infinite horizon Markov Decision Processes (MDPs) by visualizing them as deterministic processes where the states are probability measures on the original state space and the actions are stochastic kernels on the original action space. We provide a simple general algebraic approach to lifting any MDP to this space of measures; we call this to measurize the original stochastic MDP. We show that measurized MDPs are in fact a generalization of stochastic MDPs, thus the measurized framework can be deployed without loss of fidelity. Lifting an MDP can be convenient because the measurized framework enables constraints and value function approximations that are not easily available from the standard MDP setting. For instance, one can add restrictions or build approximations based on moments, quantiles, risk measures, etc. Moreover, since the measurized counterpart to any MDP is deterministic, the measurized optimality equations trade the complexity of dealing with the expected value function that appears in the stochastic optimality equations with a more complex state space.

TA43

Summit - 435

Capacity Planning for Amazon Customer Service Contact Center

Invited Session

Applied Probability Society

Chair: Xin Liu, Amazon.com Services LLC, Seattle, WA, United States

1 - Prediction interval in demand forecasting

Stephanie Mao, Amazon.com, Seattle, WA, United States

While a point forecast model can provide accurate predictions, it only represents the expected value of future realizations, failing to capture demand variability or uncertainty. Such information is crucial for businesses to plan with flexibility and meet service level targets, particularly in highly volatile situations like holidays and special events in the retail industry. For Amazon, understanding future contact volatility is further significant given its large scale, which can potentially lead to huge cost savings.

In this presentation, we introduce a new approach that generates predictive percentiles and prediction intervals to complement the given point forecast of Customer Contact Demand. This approach incorporates dynamic horizon-specific modeling and error decomposition method to balance underlying model dynamics and evolving data patterns. It prepares businesses for stochastic optimization by providing a range of potential outcomes with quantified likelihood. By leveraging prediction intervals, businesses can make more informed decisions regarding staffing, hiring, and operational cost management, especially in highly volatile demand situations.

2 - Tesseract - forecasting automation solutions

Yongfeng Hui, Amazon, Seattle, WA, United States, Chen Liang

Accurate forecasts of Customer Contact Demand play a pivotal role in the strategic planning processes of Amazon Customer Service, driving decisions on resource allocation and capacity planning. However, the dynamic nature of time series data presents significant challenges in developing effective forecasting models. Customer Contact Demands are affected by various factors, including sales events, operational changes, yearly and weekly seasonalities, and holidays. Additionally, the complexity of time series increases when covering multiple business verticals, such as countries and product lines, especially for those with limited historical data. In this presentation, we introduce an end-to-end framework tailored to address the challenges of understanding time series of Customer Contact Demand. By integrating advanced analytics with user-friendly interfaces, we ensure the accessibility of actionable insights for non-expert business users. This empowers organizations to make informed decisions, drive efficiency, and enhance customer satisfaction in the ever-evolving landscape of customer service management.

3 - A doubly stochastic Poisson arrival model for Amazon contact center: Performance evaluation and staffing implications

Kevin Melendez, Amazon.com Services LLC, Seattle, WA, United States, Xin Liu

Motivated by non-Poisson stochastic variability found in Amazon CS arrival data, we extend established service system staffing algorithms using the square-root staffing formula to allow for non-Poisson arrival processes. Using a doubly stochastic model, we aim to characterize a general arrival process that is more variable than the commonly used Poisson model. We derive the heavy-traffic limit of a many-server queueing model driven by this non-Poisson arrival process and we provide simple formulas to quantify the performance impact of the non-Poisson arrivals upon the staffing decisions, in order to achieve the desired service level goal. We conduct simulation experiments to demonstrate that the staffing algorithm is effective in achieving designated service level constraints.

4 - Short-term planning in Amazon customer service

Adolfo Rocco, Amazon, Seattle, WA, United States, Kevin Melendez

This paper introduces the short-term planning (STP) process within Amazon's customer service (CS). The STP aims to efficiently plan both internal headcount capacity and outsourced capacity to meet customer demand while adhering to labor laws, operational constraints, and service level targets. Our goal is to design the STP that optimizes the shift schedules to minimize costs and maximize operational performances, while providing the best experience for both customers and customer service agents (CSAs). To achieve this, we present three optimization models: (1) schedule generation, (2) short-term forecast allocation, and (3) shift assignment. Schedule generation determines the optimal set of shifts to cover anticipated demand. Short-term forecast allocation distributes the forecasted workload through different channels across different sites and skill groups. Finally, shift assignment matches individual agents to the generated schedules based on their skills, preferences, and operational requirements. Most of the models have been deployed into production, with an adoption level of 100% in North America, 65% in Europe, and 84% in Asian pacific region. As one of the fundamental investments, the STP initiative has standardized the shift generation and assignment across all regions. For 2024, our goal is to achieve $16MM cost savings via the CSA occupancy improvement and attrition rate reduction.

5 - Estimating interval-based tail probability of delay in real time

Ling Zhang, Amazon Corporate LLC, Bellevue, WA, United States, Yunan Liu

We develop effective methodologies and algorithms to estimate the interval tail probability of delay (iTPoD) of waiting time in an interval. The iTPoD measures the fraction of all customer arrivals in an interval [0, T] experiencing a shorter-than-w waiting time, where T is the length of the time interval and w is the waiting time threshold (e.g., T = 24 hours and w = 60 seconds). Our iTPoD predictor is dubbed the finite-interval lookahead estimator (FILE). It predicts the SL in a forward-looking fashion using input parameters including the future demand forecast, future agent capacity, average handling time, average abandonment time and real-time queue information. FILE is developed using three important building blocks: the snapshot SL estimator, the many-server fluid queueing model, and the truncated Gaussian approximations. The effectiveness of FILE has been investigated and confirmed by numerical experiments and computer simulation. We have developed a python package for FILE and it is adopted by our engineer partner in production.

TA44

Summit - 436

Simulation

Invited Session

Simulation Society

Chair: Henry Lam, Columbia University, New York, NY, United States

Co-Chair: Haoxian Chen, Columbia University, Jersey City, United States

1 - Importance Sampling Strategy for Heavy-Tailed Systems with Catastrophe Principle

Xingyu Wang, IEMS, Northwestern University, Evanston, IL, United States, Chang-Han Rhee

Large deviations theory has a long history of providing powerful machinery for designing efficient rare-event simulation techniques. However, traditional large deviations theory fails to provide useful bounds in heavy-tailed contexts, and designing efficient rare-event simulation algorithms for heavy-tailed systems has been considered challenging. Recent developments in the theory of heavy-tailed large deviations enable the design of a strongly efficient importance sampling scheme that is universally applicable to a wide range of rare events. In this talk, we provide an overview of the recent developments in the large deviations theory for heavy-tailed stochastic processes, which is followed by an account of the design principle behind the strongly efficient importance sampling scheme for such processes. The implementations of the general principle are demonstrated through examples in different contexts, including settings where the entire path of the underlying system is computationally infeasible to simulate.

2 - US Mass Killings as Stochastic Processes: Exploring the Social Contagion Hypothesis

Zachary Hornberger, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Douglas King, Sheldon Jacobson

Although mass killings in the United States are rare, these tragic events leave permanent marks on communities and evoke fear in the public. There is currently debate as to whether mass killings exhibit characteristics of social contagion. Social contagion is the phenomenon where the arrival of one event increases the likelihood of subsequent events, thus spreading behavior from person-to-person over a social network. While social contagions are frequently studied by sociologists and psychologists, operations researchers can examine the veracity of contagion claims by comparing their arrival to what would be expected from a Poisson process. When arrival data is collected in practical systems, however, it is often gathered periodically (e.g., daily, monthly, quarterly) instead of continuously due to resource or technological constraints; this is the case for available data on U.S. mass killings. Despite the prevalence of Poisson process modeling in social science, little research has explored how the periodic collection of data affects the applicability of this model. Our research studies how discretization impacts the Poisson process model's interarrival distribution and introduces the periodically-observed time-homogeneous Poisson process (PTPP) model as a viable alternative. Through the employment of the PTPP model, our findings suggest that all primary types of mass killings, when modeled as stochastic processes, behave similar to Poisson processes. This supports the conclusion that U.S. mass killings are not socially contagious.

3 - ScoreFusion: Fusing Score-based Generative Models via Kullback–Leibler Divergence Barycenters

Hao Liu, Stanford University, Stanford, CA, United States, Jose Blanchet, NIAN SI

We study the problem of fusing pre-trained (auxiliary) generative models to enhance the training of a target generative model. We propose using KL-divergence weighted barycenters as an optimal fusion mechanism, in which the barycenter weights are optimally trained to minimize a suitable loss for the target population. While computing the optimal KL-barycenter weights can be challenging, we demonstrate that this process can be efficiently executed using diffusion score training when the auxiliary generative models are also trained based on diffusion score methods. Moreover, we show that our fusion method has a dimension-free sample complexity in total variation distance provided that the auxiliary models are well fitted for their own task and the auxiliary tasks combined capture the target well. The main takeaway of our method is that if the auxiliary models are well-trained and can borrow features from each other that are present in the target, our fusion method significantly improves the training of generative models. We provide a concise computational implementation of the fusion algorithm, and validate its efficiency in the low-data regime with numerical experiments involving mixtures models and image datasets.

4 - Optimizing Input Data Collection for Ranking and Selection

Eunhye Song, Georgia Institute of Technology, Atlanta, GA, United States, Taeho Kim

We investigate a Bayesian ranking and selection (R&S) problem under input uncertainty when all solutions share a collection of independent input models. These Bayesian input models are updated as more data is collected from multiple independent sources ultimately converging to the data-generating distributions. Our goal is to design a sequential sampling algorithm that identifies the optimum under the data-generating distributions most efficiently given a data collection budget. Two types of data collections are considered: input data acquisition from each source and simulation sampling for a solution at a particular set of input models. We propose a Bayesian estimator of the optimum and characterize its probabilistic convergence rate to the optimum focusing on the case when there is a finite set of candidates for the true input distributions. We derive the optimal asymptotic static sampling ratios for the input data collection and simulation by maximizing the rate function. Based on this analysis, a sequential sampling algorithm is proposed, which is further bolstered by combining the kernel ridge regression. We benchmark our algorithm against a state-of-the-art method that consider the same R&S problem.

TA45

Summit - 437

Payment Models in Healthcare

Invited Session

Health Applications Society

Chair: Kenan Arifoglu, UCL School of Management, University College London, London, E14 5AA, United Kingdom

1 - Promoting Collaborative Care: Relative Performance-Based Payment Models for Hospitals and Post-Acute Care Providers

Hang Ren, George Mason University, Fairfax, VA, United States, Kenan Arifoglu, Tolga Tezcan

Diagnosis-Related Group (DRG) and bundled payment models are widely used in healthcare reimbursement by entities like the Centers for Medicare & Medicaid Services (CMS) and insurance companies. However, these models were primarily designed for conditions managed by a single healthcare provider in a centralized manner, often overlooking the complexities of cases requiring post-acute care (PAC) following an initial hospital stay. This can result in inadequate incentives for effective care coordination between hospitals and PAC providers, especially when treatment decisions are decentralized.

Motivated by the Comprehensive Care for Joint Replacement (CJR) payment model recently introduced by CMS, which holds hospitals accountable for the quality and cost of the entire CJR episode, including the cost of PAC, we propose simple payment models that incentivize hospitals and PAC providers to collaboratively enhance the cost efficiency and quality of care for such conditions. Our approach extends traditional payment models by introducing performance targets for all providers, encompassing the entire care episode. Using a game-theoretical model, we demonstrate that the proposed payment model elicits socially optimal actions from all providers, under various assumptions. Importantly, our models do not require detailed knowledge of the hospital-PAC network structure but rely solely on observed cost and quality outcomes within the entire system.

Furthermore, while the CJR payment model represents a positive step forward, our analysis reveals potential areas for improvement. Specifically, we suggest that holding both hospitals and PAC providers financially accountable, instead of solely focusing on hospitals, would yield further enhancements in the care delivery model.

2 - Contracting Telemedicine: A Game-Theoretic Consideration for Nhs Advice & Guidance

Zidong Liu, University of Cambridge, Cambridge, United Kingdom, Feryal Erhun, Houyuan Jiang

In the UK, General Practitioners (GPs) serve as the primary point of contact for patients, conducting initial assessments and treatments. Specialist consultations generally require a GP referral, except in emergencies. The effectiveness of this gatekeeping role is crucial for the healthcare system's success. Nonetheless, the system is plagued by diagnostic errors, referral inefficiencies, and concerns about the sustainability of traditional outpatient services. To address these issues, the Advice & Guidance (A&G) service has been introduced as an optional telemedicine gatekeeper prior to outpatient referrals. This paper explores various incentive mechanisms to maximize the effectiveness of this telemedicine service in reducing inefficiencies. We propose a game-theoretic model to evaluate the system's performance under different regulatory and contractual frameworks. Our analysis includes scenarios with no regulation, traditional fee-for-service (FFS) contracts, and two-part tariff (TPT) contracts, all of which generally fail to align system incentives. We introduce cost-sharing performance-based (COPE) contracts that incentivize both the commissioner and the hospital towards socially optimal outcomes. Furthermore, we expand our analysis to a multi-period dynamic context, allowing GPs to learn and enhance their referral quality over time. We propose dynamic cost-sharing performance-based (Dynamic COPE) contracts as a solution for long-term system coordination.

3 - Blended Payment Scheme of Nhs England: Issues and Remedies

Sara Mohammadi, UCL School of Management, London, United Kingdom, Kenan Arifoglu, Bilal Gokpinar

The National Health Service (NHS) England has implemented the blended payment scheme, integrating elements from two previously established payment systems: block contracts and Payment by Results (PbR). Block contracts entail a prospective lump-sum payment given to hospitals for accessing a defined range of services for a particular period. On the other hand, PbR is an activity-based payment under which a hospital receives a certain payment (tariff) for each admission of a patient in each Healthcare Resource Group (HRG). Under the blended approach, hospitals receive an annual fixed payment similar to block contracts to secure an adequate supply of health care services and guarantee an initial income for hospitals. Then, the fixed payment is supplemented by a variable component to reflect differences between planned and actual levels of activity. Blended payment is developed to help hospitals make the most efficient use of available resources and remain in financial balance. Consequently, it can influence hospitals' decisions about their costs and planned activity levels which reflect their preparation levels and expected capacity to deliver the anticipated level of services. Utilizing a principal-agent model, we compare hospitals' decisions regarding their planned activity levels and costs under the three defined payment schemes with socially optimal actions. Our findings reveal that due to information asymmetry between hospitals and the NHS, none of the existing payment schemes can achieve socially optimal outcomes. Consequently, we propose a new payment model to address this issue effectively.

4 - Promoting Home Dialysis for Esrd Patients: A Comparative Study of Incentive Models

Salar Ghamat, Wilfrid Laurier University, Waterloo, ON, Canada, Maryam Afzalabadi, Mojtaba Araghi

Healthcare payers are actively seeking methods to improve the quality of life for patients with end-stage renal disease (ESRD) and decrease treatment expenditures by increasing the rate of home dialysis. In this paper, we develop a game-theoretical model to study two incentive models, namely target-based and competitive, offered by a payer to encourage two heterogeneous providers' efforts in increasing home dialysis utilization for ESRD patients. We show that the payer can achieve the first-best home dialysis rates with both incentive models by carefully choosing the performance benchmarks. Our paper is the first in the incentive payment literature to model and investigate compound incentive mechanisms, including improvement and achievement benchmarks. These benchmarks encourage diverse providers to work towards achieving the award regardless of their different characteristics. Additionally, we are the first to propose a competitive incentive mechanism between healthcare providers in the healthcare operations literature. Finally, we present a numerical analysis that feeds insights into the existing practical measures of these incentive models for the payer.

TA46

Summit - 438

Optimization in Health Systems and Policy Planning

Invited Session

Health Applications Society

Chair: Anahita Khojandi, University of Tennessee, Knoxville, TN, United States

Co-Chair: Jeremy Watts, University of Tennessee, Knoxville, TN, United States

1 - Ai-Driven Management of Comorbid Hypertension and Diabetes for Shared Decision-Making

Jeremy Watts, University of Tennessee, Knoxville, TN, United States, Faezeh Bagheri, Nafiseh Payani, Denizhan Demirkol, Amy Cizik, Brian Sauer, Makoto Jones, P Jon White, Merry Ward, Andrew Farrell, Anuj Kapadia, Anahita Khojandi

Our goal is to provide clinicians with treatment alternatives for managing comorbid hypertension and diabetes that account for patient preferences to facilitate shared decision-making. To do so, we present Comorbid Hypertension AI-driven Management Pathways (CHAMP). CHAMP models comorbid hypertension and diabetes progression subject to interventions (medication and/or lifestyle changes) and estimates patient-specific outcomes through an autoregressive simulation model. Consequently, CHAMP determines a diverse set of equally good intervention policies using a reinforcement learning algorithm. These policies are assessed by quality-adjusted life years estimates, enabling the comparison of intervention policies based on combinations of medication classes (angiotensin receptor antagonists or angiotensin II receptor blockers, beta-blockers, calcium channel blockers, and diuretics) and/or lifestyle changes to accommodate patient preferences. We present CHAMP's performance for the MIMIC-IV database and provide insights.

2 - Reducing Overdose Deaths and Mitigating County Disparities Through Sequential Optimization of Substance use Treatment Center Locations

Matthew Baucum, Colorado State University, Fort Collins, CO, United States, Matt Harris, Larry Kessler, Guanyi Lu

Substance use disorder (SUD) is a pressing health concern in the U.S., and connecting communities with accessible SUD treatment is a growing public health imperative. We study the optimal allocation of substance use treatment centers across counties within U.S. states over a multi-year time horizon. We use a three-part objective function which minimizes (1) statewide overdose deaths, (2) mismatch between counties' treatment center access and overdose death rates ('inequity'), and (3) mismatch between counties' treatment center access and population ('inequality'). We also consider the temporal interdependence of counties' treatment center allocations and their death-rate-based 'equity' targets, which depend on such allocations. We use a predict-then-optimize approach based on data from over 2,500 U.S. counties between 1999 and 2016. We first estimate SUD treatment centers' causal impact on county-level overdose death rates, using SUD parity laws as an instrument. We then use these estimates in a nonlinear program that reallocates states' observed levels of SUD treatment funding. Results suggest that this approach can meaningfully reduce statewide overdose deaths and between-county disparities in treatment access. Prioritizing death rate reductions over equity/equality results in more resources being directed to high-population counties, where treatment centers are most effective. Prioritizing equity/equality in treatment access results in more resources being directed to under-served low-population counties. We also find that states differ in their Pareto-optimal tradeoffs between death rates and inequity/inequality. Finally, decision makers may wish to avoid increasing all counties' SUD funding by the same percentage, which can exacerbate between-county disparities in treatment access.

4 - Optimization of Statewide Trauma Systems Driven by Care Delivery Quality and Population Equity

Xiaonan Sun, University of Washington, Seattle, WA, United States, Shan Liu, Rebecca Maine

Organized trauma systems play a crucial role in reducing mortality among severely injured patients, ensuring timely access to critical care. However, despite their benefits, disparities persist, and measuring care quality remains complex. While many studies have focused on optimizing emergency transport services, care quality and equity have often been overlooked. In this research, we aim to develop a systematic framework to optimize a statewide trauma system based on care delivery quality and population equity.

We developed a systematic framework improving care quality while addressing population equity. This objective is achieved by establishing and assigning hospital profiles representing performance targets which can be used to guide resource allocation and operational adjustment decisions. Using state data, we established a set of comprehensive trauma care quality metrics for distinct population groups formed by sociodemographic factors and Injury Severity Score (ISS). We then created a quality index to represent trauma care quality accounting for hospital variations using a Principal Component Analysis (PCA) analysis of the quality metrics. Next, we created hospital profiles using quality index of each population group, which were estimated from data and imputed using a linear mixed-effects model. We formulated a mixed-integer linear program (MILP) to maximize the quality index of targeted population groups under various equity objectives. The model identified optimal hospital profile assignments as proxies for performance targets for the hospitals. These results help identify necessary resources for performance enhancement, guiding hospitals in making targeted improvements to better serve diverse patient populations.

TA47

Summit - 439

Advance Payment Model for Health

Invited Session

Health Applications Society

Chair: Anil Aswani, UC Berkeley, Berkeley, CA, United States

Co-Chair: Muyan Jiang, UC Berkeley, Berkeley, CA, 94706, United States

1 - Evaluating the Efficacy of Providers' Compensation Contracts in Improving Participant Retention for Clinical Studies

Xueze Song, University of Illinois Urbana-Champaign, Urbana, IL, United States, Mili Mehrotra, Tharanga Rajapakshe

To improve participant retention, the clinical study sponsors often provide monetary payments to participants and compensate providers (investigators and coordinators) for their efforts. In this work, our goal is to examine the cost performance of three widely adopted compensation contracts—fixed (FC), linear (LC), and condition linear (CLC) in improving participant retention. To this end, we analyze a sponsor's decisions regarding monetary payments and providers' efforts under a centralized model, and identify the optimal contracts for the providers under the two decentralized team structures: (i) the sponsor-investigator (SI) model and (ii) the outsourcing (OM) model.

2 - The Effect of Process Standardization in Healthcare Operations

Anand Bhatia, UNC, Chapel Hill , NC, United States

Healthcare services provided to patients with similar health conditions are known to vary. Standardization of healthcare delivery is a relatively new, yet hotly debated approach to address clinical variations. Previous research on process standardization has focused on measuring adherence to established national protocols that are available only for a limited set of disease states. Using detailed nonpublic inpatient discharge data from about 35 million inpatient stays at 296 acute care hospitals in California between 2008-2016, we create a metric that quantifies process standardization measured in terms of consistency of healthcare services rendered. We examine the impact of such process standardization on the cost, quality, and variation in quality of care delivered by a hospital. We find that process standardization is associated with a reduction in cost per discharge, readmission rates and variation in readmission rates. We also find that increasing capacity utilization and complexity of patient disease mix is associated with a reduction in process standardization, while increasing focus and process adherence is associated with an increase in process standardization. We also find that the impact of process standardization varies across different departments and different disease states, and improves cost and quality performance in majority of cases. Finally, we show that benefits of process standardization in terms of readmission rates have diminishing marginal returns at a hospital.

3 - Optimal Contract Design for End-of-Life Care Payments

Muyan Jiang, UC Berkeley , Berkeley , CA, United States

A large fraction of total healthcare expenditure occurs due to end-of-life (EOL) care, which means it is important to study the problem of more carefully incentivizing necessary versus unnecessary EOL care because this has the potential to reduce overall healthcare spending. This paper introduces a principal-agent model that integrates a mixed payment system of fee-for-service and pay-for-performance in order to analyze whether it is possible to better align healthcare provider incentives with patient outcomes and cost-efficiency in EOL care. The primary contributions are to derive optimal contracts for EOL care payments using a principal-agent framework under three separate models for the healthcare provider, where each model considers a different level of risk tolerance for the provider. We derive these optimal contracts by converting the underlying principal-agent models from a bilevel optimization problem into a single-level optimization problem that can be analytically solved. Our results are demonstrated using a simulation where an optimal contract is used to price intracranial pressure monitoring for traumatic brain injuries.

TA48

Summit - 440

Empirical Retail Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Diwas KC, Emory University, Atlanta, GA, United States

1 - Impact of Temporary Store Closures on Online Sales: Evidence from a Natural Experiment

Ragip Gurlek, Emory University, Atlanta, GA, United States, Diwas KC, Paolo Letizia

This paper examines the impact of retail store closures on omnichannel sales and consumer shopping behavior in the context of the COVID-19 pandemic. To explain the likelihood of store closure, we develop a novel instrumental variable motivated by varying geopolitical responses across the US to the pandemic.

Using data from a luxury fashion retailer, we find that when a store is closed, the volume of online orders originating from its location increases by 24%. Furthermore, when the retailer closes 10% of its stores, the omnichannel total sales (offline + online) decrease by 5.5%. Our findings indicate that the online channel enables the retailer to recover 11% of offline sales that would have otherwise been lost due to store closures. We also show that compared to existing e-shoppers, new e-shoppers are more likely to order popular product models in an effort to mitigate the mismatch risk associated with online transactions. For new e-shoppers, the likelihood of ordering a popular model stands at 70%, whereas it is 45% for existing online consumers. Additionally, the conservative behavior of favoring popular models reduces the likelihood of returns by new e-shoppers.

Even for luxury apparel requiring “touch and feel” and customer tryout, the option to purchase online proves immensely valuable. The tendency of new e-shoppers to limit mismatch risk by choosing popular products may create an opportunity for retailers to strategically target these inexperienced online customers with advertisements, product promotions, or virtual fitting rooms, all geared toward reducing online shopping risk of product mismatch.

2 - Inferring Consideration Sets from Heatmap Data

Zahra Ziaei, Frank Barton School of Business, Wichita, KS, United States, Chloe Glaeser, Adam Mersereau, Seyed Emadi, Vidyaranya Gargeya

Ignoring consideration sets in customer purchase models can lead to biased preference estimates, yet inferring these sets in brick-and-mortar contexts is challenging. We demonstrate that "heatmap data" from in-store sensors can address this issue. Unlike clickstream data in e-commerce, which tracks individuals, heatmap data shows aggregate customer traffic. Despite this, we show that heatmap data can replicate many benefits of individual-level data. In non-parametric models, heatmap data allows the identification of consideration sets that would not be possible with sales data alone. In parameterized models, it enables estimation without exclusion-restriction assumptions, reducing finite-sample bias. Using heatmap sensors in an apparel store, we obtain distinct coefficient estimates, improved predictive accuracy, and better revenue estimates for product placement decisions.

3 - Gender Pricing and Fairness in An Online Marketplace: Pink, Blue, OR Purple?

Xiaoyan Liu, Santa Clara University, Santa Clara, CA, United States, Diego Aparicio, Maxime Cohen, Davide Proserpio

Gender-based pricing disparity raises fairness concerns and can be detrimental to the economy and our society. In this paper, we investigate gender-based pricing disparity in Amazon, the major online e-commerce platform by leveraging a large dataset from Amazon.com in the year 2023. Our data set encompasses approximately 16.5 million item-price records across a wide array of product categories. We find evidence of blue tax as well as pink tax across different product categories. We also show evidence of disproportional discount patterns based on targeted genders.

4 - Green E-Commerce: Environmental Impact of Fast Delivery

Chenshan Hu, Washington University in St. Louis, Saint Louis, MO, United States, Xiaoyang Long, Jiankun Sun, Dennis Zhang

It is well-established that faster delivery in e-commerce increases consumer demand. However, the impact of faster delivery on how consumers place orders (their order frequency and basket sizes) and the subsequent environmental implications are not known. To investigate these questions, we leverage a quasi-experiment involving the opening of a new local warehouse by Alibaba Group, which led to a half-day improvement in the delivery speed for local orders. Through a difference-in-differences analysis, we find that the delivery speed improvement not only increased consumers' monthly purchasing amount by 6.70%, but also increased monthly order frequency by a higher percentage (i.e., 7.74%) and reduced the average order basket size by 0.79%. These results collectively suggest that with faster delivery, consumers purchase more on the platform but do so in more frequent and smaller orders, which implies more packaging and transportation costs for each unit of product sold. Based on these results, we conduct a detailed calibration using both public and company-specific data to estimate the increase in the platform's carbon emissions due to faster delivery. We also explore and identify two mechanisms contributing to the phenomenon: order-splitting and category expansion. We combine these insights with heterogeneous treatment effect analysis to derive managerial implications for the e-commerce platform. In particular, we find that for a platform that implements a threshold shipping policy, raising the free shipping threshold may be more effective than raising the shipping fee to reduce the environmental and operational costs associated with faster delivery.

TA49

Summit - 441

Retail Platforms and Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yiwei Wang, Zhejiang University, Haining, N/A

1 - Modeling the Influence of Unpaid Effort in Delivery Agents' Choices in Last-Mile Platforms

Lina Wang, The Pennsylvania State University, State College, PA, United States, Stanley Lim, Elliot Rabinovich

We examine decision-making dynamics in the choices that delivery agents make among jobs available in last-mile platforms. Jobs at these platforms comprise the delivery of multiple orders along routes and their compensation typically accounts for the number of deliveries and the length of the routes involved in making these deliveries (from the first to the last delivery). Rarely, however, do jobs compensate agents for the distance they must travel to make the first delivery in the routes. This distance (typically referred to as stem distance, deadhead, or empty miles) often constitutes a significant portion of the effort agents put in to complete these jobs. We seek to determine how this unpaid effort influences job selection by agents relative to other job attributes. We obtained job selection data from a last-mile delivery platform and, based on this data, used a structural framework to estimate a dynamic choice model that accounts for the effects of stem distances and other job attributes on agents' job selection decisions. This model also considers the uncertainty agents face regarding the availability and quality of future jobs by accounting for the option value of waiting as well as dynamic selections among agents. We find that an increase in stem distances markedly reduces the likelihood of job selection, overshadowing job attributes such as payout amounts and delivery lead times. Additionally we find that the information available at the platform allowing agents to calculate stem distances induces greater selectivity among agents in favor of better paid jobs.

2 - Generative AI in Action: Field Experimental Evidence on Worker Performance in E-Commerce Customer Service Operations

Lauren Lu, Dartmouth College, Hanover, NH, United States, Xiao Ni, Yiwei Wang, Tianjun Feng

We investigate how generative AI tools affect worker performance. In collaboration with an e-commerce platform, we analyze a large-scale field experiment to uncover the impact of generative AI on worker productivity and service quality in an online retail after-sales service setting. Helping after-sales service agents in interacting with customers, the adoption of the generative AI tool reduces the average problem identification time and increases the average problem resolution rate. Meanwhile, we find that the tool improves service quality as reflected in increased customer ratings and satisfaction rates.

3 - Operations Strategies of Retail Platforms: Competing Through Logistics Service

Yao Li, Kellogg, Northwestern University, Chicago, IL, United States, Lauren Lu, Jan A. Van Mieghem

We study the operations strategies of retail platforms that provide logistics service to sellers. We develop a stylized game-theoretic model where two competing retail platforms provide both the marketplace and logistics services to a single seller. The platforms first decide their commission rates and logistics service fees, and the seller decides whether to sell on the platforms. As expected, we find that in equilibrium, the commission rates decrease as the competition between the platforms intensifies. Furthermore, there exists a strategic substitution between commission rates and logistics fees.

4 - The Value of Curated Boxes: Evidence from an Online Omnichannel Fashion Retailer

Yiwei Wang, Zhejiang University, Haining, China, People's Republic of

Curated boxes offer customers a carefully chosen assortment of products, allowing them to purchase desired items and return the rest. While this channel strategy enhances customer shopping experience, its impact on retailer profitability remains largely unexplored. Employing field experiment data from an online omnichannel fashion retailer, we examine the effects of introducing and terminating a monthly curated box program on customer behaviors in an online sales channel and a subscription rental channel. We observe that curated boxes boost omnichannel total sales due to (1) direct purchases from curated boxes, and (2) demand spillovers to the online sales and subscription rental channels. During the treatment period (i.e., after introducing the curated-box program), demand spillovers to the online sales channel are less than to the subscription rental channel. During the posttreatment period (i.e., after terminating the curated-box program), demand spillovers amplify in the online sales channel but decline in the subscription rental channel.

TA50

Summit - 442

Organ Transplantation 1

Invited Session

MSOM: Healthcare

Chair: Diwakar Gupta, University of Texas, Austin, TX, United States

1 - Is the Targeted Placement of Deceased-Donor Kidneys Beneficial?

Paola Martin, Indiana University, Bloomington, IN, United States, Jingyao Huang, Diwakar Gupta

In the US, Organ Procurement Organizations (OPOs) are responsible for procuring deceased-donor organs, whereas Transplant Programs (TxPs) decide which organs to utilize for which candidates. OPOs typically make simultaneous offers for several candidates listed at a program in the strict priority sequence determined by national allocation rules. Higher-ranked candidates have the right of first refusal over lower-ranked candidates. TxPs are expected to make utilization decisions based on the merits of performing a transplant for each candidate independently of other candidates. We present evidence that this is not true in a significant number of cases. In fact, TxPs choose specific lower-ranked candidates for a subset of kidneys while skipping several higher-ranked candidates. The objective of this talk is to estimate the prevalence of such targeted placements (TPs) and quantify their impact on recipients and skipped candidates. We find that TPs have disparate impact on waitlisted candidates -- shortening the time on dialysis for TP recipients but lengthening it for skipped candidates. However, they have no significant effect on the 1-year post-transplant patient and graft survival among either group of candidates. Thus, the net benefit of TPs depends on the relative weights one would place on their positive and negative impacts.

2 - Interpretable Score Models for Predicting Discard Risk of Deceased Donor Kidneys

Osman Ozaltin, North Carolina State University, Raleigh, NC, United States, Ruoting Li, Sait Tunc, Matthew Ellis

Despite their growing shortage, more than 20% of the donated kidneys are discarded, a substantial proportion of which is due to organ quality and matching this quality to a recipient's needs. Therefore, mechanisms to expeditiously match available organs at higher risk for discard with patients that may benefit from these organs are critical. This study aims to predict the discard risk of deceased donor kidneys using interpretable models. We propose risk scores models using two approaches. First, we develop a discard risk score without using KDRI. In the second approach, we extend KDRI to estimate discard risk. We present a computational evaluation of prediction performance and drive key insights into factors affecting discard risk.

3 - The Impact of Competitive Entry in An Organ Transplant Market

Guihua Wang, The University of Texas at Dallas, Richardson, TX, United States, Jiayi Liu, Jun Li, Wallace Hopp

Using patient-level data from all centers for organ transplant in US, we provide empirical evidence that new entrants affect incumbents' risk-taking behaviors. We then develop an econometric model to analyze the impact of competitive entry on patients' waiting time and survival.

TA51

Summit - 443

Digital Marketplaces and Platforms, and Online Advertising

Invited Session

MSOM: Service Operations

Chair: Sami Najafi Asadolahi, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Naren Agrawal, Santa Clara University, Santa Clara, CA, United States

1 - Double Distributionally Robust Bid Shading for First Price Auctions

Yanlin Qu, Stanford University, Stanford, CA, United States, Ravi Kant, Yan Chen, Brendan Kitts, San Gultekin, Aaron Flores, Jose Blanchet

Bid shading has become a standard practice in the digital advertising industry, in which most auctions for advertising (ad) opportunities are of first price type. Given an ad opportunity, performing bid shading requires estimating not only the value of the opportunity but also the distribution of the highest bid from competitors (i.e. the competitive landscape). Since these two estimates tend to be very noisy in practice, first-price auction participants need a bid shading policy that is robust against relatively significant estimation errors. In this work, we provide a max-min formulation in which we maximize the surplus against an adversary that chooses a distribution both for the value and the competitive landscape each from a Kullback-Leibler-based ambiguity set. As we demonstrate, the two ambiguity sets are essential to adjust the shape of the bid-shading policy in a principled way to effectively cope with uncertainty. Our distributionally robust bid shading policy remains efficient to compute and systematically outperforms its non-robust counterpart on real datasets provided by Yahoo DSP.

2 - Dynamic Pricing and Bidding for Display Advertising Campaigns

Sami Najafi Asadolahi, Santa Clara University, Santa Clara, CA, United States, Naren Agrawal, Stephen Smith

Display advertising’s growth has transformed business engagement with target audiences. Despite its benefits, managing campaigns is challenging due to uncertainties in arrival rates, user visits, and real-time auction outcomes. We aim to devise an optimization method for agencies to profitably price and bid for campaigns while maintaining clients’ expectations. We formulate real-time bidding strategies for impressions or specific actions performed by viewers (e.g., clicks) and pricing strategies for new campaigns, through a Markov decision process to maximize profits. Through analysis, we offer valuable insights, demonstrating the efficacy of strategies based on campaign arrival intensity. We identify an optimal approach to allocating viewers to different campaign types based on their bids for viewers. Conceptually, this approach visualizes an internal market operated by the ad agency, where the various campaign types compete for viewers. To facilitate the implementation of our methodology, we also develop heuristic solution methods that can be scaled to accommodate the large numbers of viewer impressions or actions often requested in practice. We quantify the operational benefits resulting from combining campaigns that target similar viewer types, or merging similar ad agencies. This leads to higher profits for the ad agency, lower optimal prices, and reduced delays in delivering promised impressions or actions to campaigns. We present a practical, easily implementable heuristic that efficiently consolidates requested viewer impressions or actions into larger blocks, significantly reducing computation time while maintaining minimal impact on expected profit.

3 - Optimal Provision of Data via Collaborative Learning Platforms Under Competition

Mingxi Zhu, Georgia Institute of Technology, Atlanta, GA, United States

This study explores the merchants optimal provision of data samples via a collaborative learning platform, realizing the future potential competitive dynamics after data sharing. With the increasing recognition of data as a strategic asset, firms are confronted with the decision of whether to share their data with competitors, thereby potentially compromising their competitive advantage, or withhold it to maintain exclusivity at the expense of potential collaborative benefits. This work provides a novel framework to first explore the firms optimal decision on data sharing without platform intervention. Then, it studies how platform mechanism design on data sharing would induce a Pareto improvement among firms. The results presented in this paper shed lights in how to balance between the advantages of cooperative data sharing and the risks associated with divulging additional information to competitors, and provides insights for firms seeking to optimize their data sharing decisions in competitive markets.

TA52

Summit - 444

Retail Analytics: Space and Assortment Planning

Invited Session

MSOM: Service Operations

Chair: Tulay Flamand, University of Colorado Denver, Denver, CO, United States

1 - Store-Wide Shelf Space Allocation under Health Safety Considerations

Tulay Flamand, University of Colorado Denver, Denver, CO, United States, Ahmed Ghoniem, Bacel Maddah

Given the increased public health concerns stemming from events like the COVID-19 pandemic, managing the health safety of public spaces such as grocery stores has become crucial. This study addresses a store-wide shelf-space allocation problem in grocery stores with a focus on health safety. We propose a mixed-integer nonlinear programming model that strategically allocates grouped product categories to shelves. This allocation guides in-store traffic and minimizes peak traffic density across all shelves, thereby reducing customer interaction and enhancing safety. We present our preliminary results to demonstrate the efficacy of our model.

2 - Pricing And Convenience Considerations In Omnichannel Assortment Planning

Joshua Gladstone, University of Massachusetts Amherst, Amherst, MA, United States

We address the problem of assortment planning and pricing optimization for omnichannel grocery retailers. In the proposed mixed-integer nonlinear program, the decision-maker jointly optimizes channel assortments and prices, endogenously driven by consumer price and convenience preferences, while accounting for limited brick-and-mortar shelf space and economies of scale incentives for the retailer. Linearizing our model enables computational analysis and provides managerial insights for simultaneously optimizing online and in-store product lines.

3 - Bi-objective Optimization for Store-Wide Shelf Space Allocation

Bacel Maddah, American University of Beirut, Beirut, Lebanon

We develop a framework for optimizing shelf-space allocation in grocery stores that balances impulse buying with shopping convenience. Our model, a bi-objective nonlinear integer program, is applied to a large supermarket in France. We propose a linearization scheme, and obtain the efficient frontier of Pareto-optimal solutions. Results show potential for significantly boosting impulse revenue while reducing customer inconvenience.

4 - Inventory control with FIFO and LIFO picking behavior

Anna-Lena Sachs, Lancaster University, Lancaster, United Kingdom, Thomas Vogt, Ulrich Thonemann, Ben Lowery

Customer picking behaviour plays an important role in retail inventory management. Standard inventory models usually distinguish between picking the newest items, i.e. Last-in-first-out (LIFO), or the oldest items first, i.e., First-in-first-out (FIFO). We analyze how LIFO and FIFO picking behaviour affects inventory management in retailing, and whether monetary and non-monetary incentives can change customer picking behaviour to increase sales of earlier expiring items and reduce food waste in retailing.

5 - Who gets the product? Channel Assortment and Inventory Allocation Strategy Design for an Omnichannel Retailer

Jia Guo, California State University-East Bay, Dublin, CA, United States, Iman Dayarian, Burcu Keskin

We explore the channel assortment and inventory allocation decisions for a dual-channel retailer that operates both physical and online stores. The retailer primarily fulfills online orders through distribution center deliveries and employs omnichannel strategies like Ship-from-Store and Buy-Online-Pickup-in-Store, leveraging in-store inventory to fulfill online orders. Despite competitive pressure to expand fulfillment channels, operational challenges increase fulfillment costs and decision complexity. Customers with different preferences on fulfillment channels are modeled through customer types. Each customer type corresponds to a specific preference order among fulfillment channels, with customers purchasing from their highest-ranked available channel. The retailer's challenge involves solving a channel assortment problem and determining which channels to offer customers. Simultaneously, inventory allocation decisions entail choosing a subset of the channel assortment for every sequentially arrived customer during the selling season to maximize expected profit. Our analysis reveals that adding fulfillment channels through omnichannel strategies can reduce profits due to customers' stockout substitution behavior, highlighting the role of demand information and dynamic rationing in mitigating these effects and maximizing the benefits of implementing omnichannel strategies. Dynamic rationing allows retailers to adjust their decisions in response to real-time sales data, providing flexibility and adaptability. Our findings suggest that a dynamic rationing policy, developed with limited demand information, may outperform a static allocation policy based on precise demand information in specific scenarios.

TA53

Summit - 445

Supply Chain Finance and Fintech

Invited Session

MSOM: Supply Chain

Chair: WENTING LI, Arizona State University, Tempe, AZ, 85044-1746, United States

Co-Chair: Rafael Escamilla, Arizona State University, Tempe, AZ, United States

Co-Chair: Rafael Escamilla

1 - Text Beyond Text: How Short Videos' Sentiment Impact Stock Market

Yifan Ren, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Qiyu Dai

With the growing availability of financial news on short video platforms, the sentiment in the videos is affecting investors' attitudes towards the performance of firms. Such an impact can be reflected in the return of stocks. In this paper, we aim to examine the impact of short videos' sentiment on the future return of stocks for a rich set of companies. We extract the texts in all videos posted by some official financial news accounts. Based on the tag of videos, we match each video to firms it discusses. Then, each videos' anticipation towards the future trend of stocks as either "Long" or "Short" based on their textual information using with the help of large language models. A comparison of the stock return between the two groups reveals that companies that are categorized as "Long" have slightly higher future returns, highlighting the importance of short video financial news on the stock market. Our study provides managerial implications for both stock investors and listed companies.

2 - Supply Chain Finance and Payment Practices: a Dyadic Perspective

Yiyi Fan, Lancaster University, Lancaster, United Kingdom, Fang Li, Mark Stevenson, Shantanu Banerjee

This research examines the effect of supply chain finance (SCF) on payment practices from a dyadic, buyer-supplier perspective. We exploit current UK policy, which mandates that larger businesses must disclose their payment practices, including whether they make SCF available to their suppliers, to obtain a buyer-supplier matched sample. We examine whether using SCF offered by disclosing buyer firms affects their payment practices, including late payments to suppliers. We find that SCF is positively associated with the average payment time of buyers and with a lower proportion of free-of-charge fast payments, suggesting SCF is a trade-off decision for suppliers between speed of payment receipt and cost. Moreover, firms offering SCF have a higher proportion of late payments, i.e. invoices that are not paid within the agreed payment period, and are more likely to negotiate a longer contractual payment period with their suppliers. We also find that using SCF is negatively associated with suppliers' receivable days, suggesting suppliers could indeed obtain faster payments through SCF but must bear the cost. Collectively, our results suggest that SCF initiated by large buyer firms may speed up payment receipt for suppliers in general, but the problem of late payments becomes worse for suppliers unless they pay a premium to receive what they are owed. The findings challenge the prevailing assumption that SCF is beneficial and point to the potential risk of creating more adversarial buyer-supplier relationships.

3 - Supply-Chain Finance: An Empirical Evaluation of Supplier Outcomes

Yingjie Qi, Copenhagen Business School, Copenhagen, Denmark, Niklas Amberg, Tor Jacobson

Buyers and suppliers have diverging interests about trade-credit maturities: buyers desire long payment periods as a source of cheap funding, while suppliers prefer swift payments to avoid locking up scarce liquidity in idle assets. A fast-growing financial product innovation---supply-chain finance (SCF)---offers to resolve these diverging interests, but its net effect on suppliers is a priori unclear. We study the effects of SCF programs on suppliers using unique invoice-level data from a large Swedish bank. We find that SCF programs relax suppliers' liquidity constraints and thereby enable them to grow their sales, employment, and investments.

TA54

Summit - 446

Advances in Operations Management

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Eryn Juan He, University of Utah, Salt Lake City, UT, United States

1 - FinTech Lending, Open Banking, and Consumer Manipulation

Yangguang Wang, Syracuse University, Syracuse, NY, United States, Zhengping Wu, Fasheng Xu

Ensuring equality of opportunity is pivotal for personal development. Nevertheless, roughly 3 billion disadvantaged individuals worldwide are hindered from realizing their full potential due to inadequate access to financial services. Over the last decade, fintech companies have been at the forefront of addressing this issue by leveraging alternative data for credit scoring. Despite notable progress, challenges persist, particularly regarding the manipulation of alternative data by individuals. Against this backdrop, Open Banking is emerging on a global scale, with the goal of enabling the free sharing of banking data and thus promoting financial inclusion. To examine the effects of Open Banking on fintech companies' services for disadvantaged populations, we introduce a game-theoretical model wherein individuals strategically decide whether to manipulate their alternative data to secure a loan from fintech lenders. Although Open Banking enables fintechs to access individuals' additional data source – banking data, we raise a main concern. Making such banking data available to fintech lenders may alter consumers' manipulative behavior in a way that dilutes the information contained in the original alternative data source, thereby making the fintech less efficient at evaluating individuals' default risks. The information asymmetry between these two parties may become more severe, leading to a more conservative lending practice of the fintech lenders, and ultimately weakening financial inclusion.

2 - A Framework for New Venture Creation

Zhengli Wang, The University of Hong Kong, Hong Kong, Hong Kong, Stefanos Zenios

We model the creation of a new venture with a diffusion control framework. The state of the venture is captured by a diffusion process. The entrepreneur chooses between two controls that determine the drift and the variance of the process. Depending on whether the process reaches the upper or the lower boundary, the venture succeeds or fails. The entrepreneur wishes to maximizes the expected total payoff. We derive the optimal policy and show that the gap between using the optimal policy and using the best single control is small.

3 - Optimal Contract Design under Multiple Moral Hazard - Ex Ante, Ex Post Cash Diversion, and Price Deviation

Ruiting Zuo, Financial Technology Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, People's Republic of, Mingliu Chen, Feng Tian

We consider a decentralized newsvendor production problem subject to agency issues. A principal funds and contracts with an agent for production planning. However, the agent is subject to moral hazard, such as cash diversion during different stages of a production process, which is unobservable to the firm. We derive the principal's optimal contract and corresponding operational decisions under ex ante and ex post cash diversion scenarios, which happen before and after demand is realized, respectively. When only subjected to ex ante diversion, the optimal contract is a debt contract, which leads to underproduction and a higher price than that under first-best centralized production. When subjected to both ex ante and ex post diversion, the debt contract is no longer optimal. Instead, we present a modified debt repayment that resembles both debt and revenue-sharing contracts, which is optimal. Furthermore, we find that the optimal capacity and price depend on the level of ex post diversion efficiency with respect to which the optimal price is nonmonotonic. Finally, we consider a scenario where the actual sales price is also unobservable and find that the principal can still use a contract similar to debt to deter ex ante diversion and price deviation in most regimes. However, this contract must have an additional clause penalizing excessive revenue. All optimal contracts in this paper have simple forms and are easy to implement. We highlight the optimal operational decisions, such as the optimal capacity and price, accompanied by the financial decisions.

4 - Ethics and Platform Responsibility

Nabita Penmetsa, University of Utah, Salt Lake City, UT, United States, Manu Goyal, Krishnan Anand

Should platforms voluntarily assume responsibility for losses incurred by consumers in the absence of legal mandates?. To answer this question, we develop a finite-horizon model of a platform market with the platform, agents, and consumers. The platform is either ethical or unconstrained and is called into action whenever a consumer incurs a loss. The agents are short-run and control both the choice of effort and prices. Consumers are strategic and play a critical role in revealing the platform type by posting reviews. We show that an ethical platform, one that assumes responsibility voluntarily by compensating the consumer for losses, outperforms a platform unconstrained by such ethics. Moreover, the mere possibility of ethical platforms in the population can increase effort and prices over time.

We show that without reviews, even an ethical platform cannot induce agents to exert effort, and prices remain forever low. However, when consumers post reviews with some probability, however small, even an unconstrained platform can induce agents to exert effort and cause prices to increase over the horizon. When number of periods is large, it is indeed optimal for the unconstrained platform to assume responsibility and protect consumers from losses.

5 - Dissecting the Learning Curve for Supply Chain Efficiency and Stability: An Empirical Analysis

Mengfei Li, Fudan University, Shanghai, China, People's Republic of, Yue Cheng, Liu Ming, Xiaole Wu, Wenchang Zhang

Our study explores learning effects in the supply chain and production processes of small/median sized manufacturing firms(SMEs) on both efficiency and stability. We focus on the enhancement of four key supply chain stages - upstream supply chain, scheduling, production, and downstream supply chain - through learning, and the impact of learning materials(historical orders) factors and spillover effect on learning effectiveness. The research is grounded in a rich dataset from dozens of SMEs manufacturing firms, allowing for precise tracking of each production process's timeline and various pieces of information.We employ a range of empirical strategies to analyze how production efficiency and stability evolve as firms' experience accrues. We further reveal underlying mechanisms by examining the learning effects of orders' batch sizes and urgency, as well as potential heterogeneous spillovers among different product lines. These insights offer practical recommendations for managerial optimization, leading to improved organizational effectiveness, and underscore the vital role of continuous learning in enhancing manufacturing efficiency and stability, a critical determinant of the competitiveness of order-responsive enterprises.

TA55

Summit - 447

Workforce Development and Operations Research Models

Invited Session

Public Sector OR

Chair: Guanting Wu, 15289

Co-Chair: Peter Zhang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Labor Implications of the Energy Transition: Assessing Transforming and Emerging Industries

Jillian Miles, Carnegie Mellon Unviersity, Pittsburgh, PA, United States

Amid global efforts to mitigate climate change, large industries are facing increasing pressure to decarbonize, particularly in heavy manufacturing sectors such as iron and steelmaking, while emerging 'green' industries, like battery manufacturing for electric vehicles, are working towards greater prominence. However, these technological shifts, driven by the broader energy transition, could profoundly affect local labor markets. This analysis presents two case studies that integrate industry and jobs data with occupational characteristics to evaluate how technological advancements or growth may impact local labor dynamics. The first case examines the iron and steelmaking industry, focusing on the shift from high-emitting blast furnace "integrated" production to low-emitting electric arc furnace (EAF) processes. Our findings indicate that while traditional iron and steel workers possess the necessary skills for EAF production, less than half of the original workforce is required to sustain the new operations, putting a significant number of production workers at risk of unemployment. The second case study explores the labor readiness of various U.S. locales considering incentivizing battery manufacturing, by measuring the possible shock adding a small or large battery gigafactory would have on a given area. Preliminary results show varied readiness for both sizes of factories and explores overlaps with disadvantaged communities and energy communities. Both studies provide critical insights for policymakers, highlighting the potential need for skills training and wage insurance, and underscore the importance of centering community needs in the analysis.

2 - Learning by Doing Different Gigs: Uncovering Latent Skill Development on Online Labor Platform

Guanting Wu, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang, Hai Wang

Problem definition: Online labor markets (OLMs) are experiencing rapid growth, in which online labor platforms help clients and workers find each other based on relevant skills necessary to complete projects. Their online interactions reveal the learning trajectories of individual workers by doing different gigs. In this paper, we uncover how workers develop skills and how their diverse skills interdepend on online labor platforms. Methodology: We propose a partially observable Markov process (POMP) framework and instantiate it to a parameterized multivariate linear-Gaussian model. We also identify a sampling selection bias in discrete-time data, which motivates model design. Due to the intrinsic data sparsity, we estimate the parameters of the multivariate model using a two-stage solution approach. Results: Using a proprietary and large-scale dataset containing millions of transactions over a 12-month period from a leading online labor platform, we apply our method to gain insights into the skill development of workers and the interdependence of diverse skills. First, workers develop their skills through the completion of increasingly complex projects. Second, we observe a significant interdependence between project and skill categories. Third, project complexity negatively affects service quality. Managerial implications: Based on our findings, we propose that online labor platforms could design growth-aware recommendation systems and broaden the criteria for worker recommendations considering their diverse skills. In addition, platforms could encourage clients to decompose projects into several sub-projects to reduce project complexity.

3 - Robust contract design and linear contract optimality framework

Michelle Wang, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang

tbd

4 - Analyzing the Gig Economy via Agent-Based Modeling with a Focus on Gig Workers Challenges

Tara Eslaminokandeh, Northeastern University, Boston, MA, United States, Kunind Sharma, Ozlem Ergun, Michael Kane

With the expansion of the gig economy and the use of digital platforms in recent years, sectors such as transportation and accommodation are highly impacted. Although this growth offers numerous advantages, it also presents considerable drawbacks that must be addressed and resolved. Most studies in this area tend to focus on the perspectives of platforms and customers, often overlooking the interests and challenges faced by workers, such as information asymmetry, which negatively impacts gig workers and complicates their decision-making processes.

In this study, we analyze the gig economy by focusing on addressing the critical challenges and concerns from the perspectives of gig workers. Specifically, we employ an agent-based modeling (ABM) method to simulate the behaviors and interactions of stakeholders in the gig economy—workers, customers, and platforms—to study the impact of these interactions on both the system and the individuals involved. Our methodology comprises three phases: collecting data through surveys and interviews, developing an ABM simulation model based on this data, and designing and conducting experiments to analyze various scenarios within the gig market. We particularly examine worker responses to different policies and systemic changes. This approach contributes to a comprehensive understanding of the gig economy, emphasizing the importance of addressing worker concerns and developing equitable policies to support a more sustainable gig economy.

Our findings aim to inform policymakers and platform designers about the effects of their decisions on all stakeholders, helping them to ensure that the gig economy benefits workers, customers, platform companies, and society simultaneously.

TA56

Summit - 448

Modeling and Optimization: Shaping an Equitable Future

Panel Session

Minority Issues Forum

Co-Chair: Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara

Co-Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, 95053, United States

1 - Moderator Panelist

Xiaoyan Liu, Santa Clara University, Santa Clara, CA, United States

2 - Panelist

Samantha Keppler, University of Michigan, Ann Arbor, MI, United States

3 - Panelist

Alfonso Pedraza-Martinez, University of Notre Dame, Notre Dame, IN, United States

4 - Panelist

Telesilla Kotsi, The Ohio State University, Columbus, OH, United States

5 - Panelist

Maggie Zhang, Santa Clara University, Santa Clara, CA, United States

TA57

Summit - Terrace Suite 1

Equity in Health Decision-Making: Navigating the Future Together

Invited Session

Health Applications Society

Chair: Gian-Gabriel Garcia, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Amaya McNealey, Georgia Institutue of Technology, Atlanta, GA, 30318, United States

1 - Race-Neutral Calculator Still Disadvantages Black Patients

Amaya McNealey, Georgia Institute of Technology, Madison, AL, United States

There has been a longstanding concern in healthcare about the appropriateness of using race and ethnicity in predictive algorithms to facilitate prognosis, diagnosis, and treatment. Specifically, the VBAC calculator is a clinical decision support tool to identify if patients who have previously experienced a low-transverse cesarean delivery are eligible to be recommended a trial of labor after cesarean (TOLAC) or if they are recommended a repeat cesarean delivery. The original VBAC calculator created in 2001, “VBAC 1.0”, faced criticism after explicitly using race/ethnicity as a variable within the model. To replace VBAC 1.0, “VBAC 2.0” was created in 2021, removing the use of race/ethnicity and adding an indication of treated chronic hypertension. However, while the statistical performance of VBAC 2.0 was similar to VBAC 1.0, it is not explicit to what extent VBAC 2.0 shifts risk scores for racial groups and subsequently how these updated predictions affect the likelihood of recommending a TOLAC. To compare the effects of both calculators, we conducted a secondary analysis of pregnant patients who had a prior low transverse delivery at Grady Memorial Hospital in Atlanta, GA. Our results show that 34.0% of Non-Hispanic Black patients would have been recommended a cesarean delivery due to the VBAC-1.0 compared to 60.7% under VBAC-2.0 (Likelihood ratio (LR): 1.78; 95% CI: 1.52-2.10). For non-Hispanic Black patients, the increased likelihood of a cesarean recommendation remained regardless of their respective hypertension status, indicating that additional research in machine learning fairness is crucial to ensure equitable shared decision-making and downstream effects.

2 - Fairness-Promoting Stochastic Facility Location

Karmel Shehadeh, Lehigh University, Bethlehem, PA, United States

Uneven distributions of healthcare facilities across different regions lead to disparities in accessibility and transportation distance to health services. To mitigate such disparities, we propose fairness-promoting stochastic optimization methodologies for the stochastic fixed-charge facility location problem. Our results show, for the first time, how different methods of modeling fairness under uncertainty can result in different location decisions with varying impacts on fairness. In addition, they underscore the potential of the proposed approaches to enhance fairness in facility location.

3 - The Utility and Fairness of The Inclusion of Race in Cardiovascular Risk Prediction Algorithms

Noah Hammarlund, University of Florida, Gainesville, FL, United States

Cardiovascular disease (CVD) remains a leading cause of mortality worldwide, with risk prediction algorithms playing a pivotal role in risk assessment and intervention by matching prevention efforts to individual needs. Recent debates have surfaced regarding the exclusion of race as a variable in these CVD prediction algorithms due to concerns about perpetuating biological racism in clinical practice. Concurrently, there is growing interest in integrating social determinants of health (SDoH) into prediction models in place of race. However, the impact of excluding race on the utility and fairness of CVD prediction models remains uncertain, prompting the need for a critical reevaluation. This study investigates utility and fairness considerations surrounding the exclusion of race and the potential integration of SDoH in cardiovascular risk prediction models.

Using the CARDIA (Coronary Artery Risk Development in Young Adults) dataset, this study compares the predictive efficacy of CVD risk prediction algorithms with and without race as a predictor. The study also explores the ethical and fairness implications of SDoH-informed algorithms by comparing models with SDoH predictors to traditional models with race. Given recent concerns about the fairness of race-based risk assessments, the research seeks to determine whether the replacement of race with SDoH predictors may alleviate or exacerbate existing health disparities.

Through a comprehensive analysis, this study aims to inform evidence-based decision-making in cardiovascular medicine. By leveraging insights from the CARDIA dataset, it strives to develop ethically sound recommendations for risk prediction models that balance accuracy and equity, ultimately improving health outcomes for diverse populations.

4 - Personalized Visit Planning for Patients with Chronic Disease

Katherine Adams, University of Wisconsin-Madison, Madison, WI, United States, Justin Boutilier, Sarang Deo, Yonatan Mintz

Approximately 240 million people are living with undiagnosed diabetes globally. Among these, almost 90% reside in low- and middle-income countries (LMICs), leading to a large proportion of people with unmanaged blood glucose levels. Effective interventions require the joint planning of screening and management visits. Community Health Worker (CHW) programs provide an affordable and culturally tailored solution for early detection and management of diabetes whilst expanding health system capacity. We propose an optimization framework to personalize CHW visits, tradeoff screening and managing patients, and maximize glycemic control. By modeling patient motivational states, we also anticipate and prevent drop out decisions for enrolled patients. We present structural and computational results using real data from India.

5 - Addressing Subgroup-Specific Bias Using Multi-Level Approaches and Causal Inference

Vishwali Mhasawade, New York University, New York, NY, United States

While a patient visits the hospital for treatment, factors outside the hospital, such as where the individual resides and what educational and vocational opportunities are present, play a vital role in the patient's health trajectory. On the contrary, most advances in machine learning in healthcare are mainly restricted to data within hospitals and clinics. Moreover, data is limited chiefly to observed factors assumed to be proxies for the quantities of interest. While health equity, defined as minimizing avoidable disparities in health and its determinants between groups of people with different social privileges in terms of power, wealth, and prestige, is the primary principle underlying public health research, this has been largely ignored by the current machine learning systems. Inequality at the social level is harmful to the population as a whole. Thus, focusing on the factors related to health outside the hospital is imperative to address specific challenges for high-risk individuals and determine what policies will benefit the community as a whole. In this talk, I will first demonstrate the challenges of mitigating health disparities resulting from the different representations of demographic groups based on attributes like gender and self-reported race. Next, I will present a causal remedial approach to health inequity using algorithmic fairness that reduces health disparities. Ultimately, I will discuss how the appropriate choice of labels used for training machine learning algorithms can impact fairness considerations.

TA58

Summit - Terrace Suite 2

OR Applications for Medical Decision-making

Invited Session

Health Applications Society

Chair: Daniel Otero-Leon, Harvard Medical School / University of Virginia, Boston, MA, United States

1 - Evaluating Disparities in Colorectal Cancer (Crc)

Bryan Ortiz Torres, University of Wisconsin Madison, Madison, WI, United States, Gabriel Zayas-Caban

Health disparities in colorectal cancer (CRC) patients have been widely documented. While CRC incidence rates have decreased during the past decades, differences between non-Hispanic black and non-Hispanic white individuals still exists. Here, we present our work on analyzing disparities for CRC patients using survival analysis and data from the Southern Community Cohort Study (SCCS).

2 - Clustered Multi-Task Learning for Prediction of Adverse Pregnancy Outcomes

Sun Ju Lee, Georgia Institute of Technology, Atlanta, GA, United States, Gian Garcia, Kaitlyn Stanhope, Marissa Platner, Sheree Boulet

In clinical prediction models, combining individual indicators into a composite outcome can help overcome issues associated with single outcomes such as low predictability and low prevalence. However, this may obscure relationships between predictors and single outcomes, and may limit the clinical utility of the prediction model as diagnoses may have different etiologies addressed by different interventions. We aim to resolve this trade-off by developing an optimization framework to simultaneously cluster related outcomes and learn model parameters for each cluster. We apply our formulation to indicators comprising maternal and neonatal morbidity and demonstrate that our approach can aid interpretability by finding underlying groups of related tasks and deriving an interpretable set of predictors.

3 - Analyzing the Dynamics of Overdose Death Spikes by Substance Across Massachusetts Counties

Daniel Otero Leon, Harvard Medical School, Boston, MA, United States, Mohammad Jalali, Huiru Dong, Hannah Lee

Overdose death data trend analysis often consists of yearly aggregate trends or rates, which may mask short-term fluctuations. Emergency responders often report that periods of calm are followed by a sudden rise, or spikes, in overdoses, which suggests that the risk of death can change substantially in a short period. Multiple substances are involved in overdose death, and it will be crucial to know if certain substances increase the likelihood of a spike. This study aimed to analyze at a county level in Massachusetts which substances are involved more during spike days. We count with overdose fatalities from the Massachusetts Registry of Vital Records and Statistics from 2017 to 2022. We employed a peak detection algorithm based on z-scores that identifies deviations from the established patterns within a dataset. We then applied t-tests in select counties to compare substance prevalence between spike and non-spike days.

TA60

Summit - Ballroom 2

Amazon Warehouse Resource Optimization

Invited Session

The Practice Section of INFORMS

Chair: Xinyu Fei, Amazon.com, Bellevue, WA, 98004, United States

1 - Labor Staffing in Amazon Sortable Fulfillment Centers

Yingqiu Zhang, Amazon, BELLEVUE, WA, United States

When customers place the order through Amazon.com, associates in Fulfillment Centers (FC) will prepare the package for shipments. There are hundreds of associates in an FC works in various processes to ensure the orders are delivered to customers' doorsteps on time. Amazon FCs are powered by technologies to effectively staff associates to processes by optimizing associate experience, reducing associate injuries, lowering cost, and maximizing the volume throughput. Picking is a critical department in ARS FCs that holds the inventory storage and allows associates to pick the right inventory items to fulfill shipments. Associates are staffed stationary at picking stations, pick required items from the inventory pods moved by robotic drives, and place picked items into totes that will be sent to downstream for further processing.Staffing problem is complex at Amazon. Due to FC's physical layout, different picking stations have different priority, and there are multiple different types of picking stations, which have their unique staffing rule. Moreover, Amazon values associates health and perform safety rotations to reduce injuries and Amazon also respects associates' accommodations for best associates' experience. In this presentation, we present an optimization model to effective staff associates in FC, and more specifically, in picking department. The problem is formulated as mixed-integer goal programming that respects business and associates' rules and optimizes picking floor performance in business priority. This model has been launched across Amazon FCs by April 2024 and improved picking floor throughput and associates experience.

2 - Real-TIME Staffing Plan Adjustment in Fulfillment Centers

Xinyu Fei, Amazon, Bellevue, WA, United States, Andrew Johnson

Amazon's fulfillment centers (FCs) play a crucial role in ensuring efficient and timely delivery of orders placed by customers. These state-of-the-art facilities employ hundreds of associates working across various processes to prepare packages for shipment. Effective staffing in these fulfillment centers directly impacts associates' working experience, operational costs, and delivery time. One of the most important systems in Amazon FCs is the multi-products packing system. Because workers have varying levels of efficiency and conditions in the system can change, adjusting worker staffing in real-time is highly required. We build a non-linear mixed-integer optimization model that considers the multi-products packing system status and business needs to optimize worker staffing. Furthermore, we've proposed a heuristic method to provide a real-time staffing solution to improve user acceptance in practice.

3 - Same Day Outbound Resource Planning

AARON HERMAN, Amazon, Seattle, WA, United States

A high quality and accurate resource plan is a crucial component for a successful shift in an Amazon Fulfillment Center (FC). Without a high quality plan, resources will not be properly utilized to process shipments in time for the customer in the most efficient manner. Without an accurate plan, other systems or warehouse operators themselves will not have the information they need to identify a capacity constraint and actions to correct for it. In this presentation we will go over the main ideas and challenges behind producing these plans for Amazon FCs.

4 - Outbound Flow Management Systems

Maximilian Zellner, Amazon, Santa Monica, CA, United States

Amazon.com continues to improve delivery speed and to reduce cost. One key initiative of Amazon Fulfillment Technology contributing to these goals is the improvement of flow between processes in the outbound area of fulfillment centers. In this area, picking and packing processes are separated by conveyance with uncertain cycle times, and the downstream processes constitute the system's bottlenecks. To achieve good flow and low costs, a control model should aim to achieve high utilization of the bottleneck resource while keeping WIP low to produce favorable cycle times. We propose a linear programming model that explicitly models flow between the picking and packing processes, and controls the amount of work at the bottleneck process by setting upstream pick throughputs.

5 - Modeling Flow in Fulfillment Centers

Andrew Johnson, Amazon, Bellevue, WA, United States, Maz Zellner

As Amazon continues to increase customer delivery speeds, modeling flow within the fulfillment center is critical. Amazon Fulfillment Technology is redesigning the flow management service to model explicitly buffer targets and cycle time within the fulfillment center. The flow management service models picking, the upstream process, that feeds multiple downstream packing processes. A critical input to the new service is the estimates for minimum, target, and maximum buffer sizes. The minimum buffers are calculated by estimating units per hour as a function of the buffer size. While a copula model is used to estimate the target buffer size. The maximum buffer sizes are calculated based on historical observed maximum buffer sizes.

TA61

Summit - Ballroom 3

What Does Our Past Tell Us About the Future

Panel Session

Committee's Choice

Co-Chair: Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

1 - Moderator Panelist

Christopher Ryan, University of British Columbia, Vancouver, BC, Canada

2 - Panelist

Linus Schrage, LINDO Systems, Inc., Chicago, IL, United States

3 - Panelist

Anna Nagurney, University of Massachusetts Amherst, Amherst, MA, United States

4 - Panelist

Thomas Magnanti, MIT, Cambridge, MA, United States

TA62

Summit - Signature Room

Social Media Dynamics and Data Analysis

Contributed Session

Chair: Anna Gao, Philadelphia, PA, United States

1 - Dynamics of sentimental involvement on social media during crisis events: case studies in China

boyu zhu, Shanghai Maritime University, Shanghai, China, People's Republic of

This study explores the dynamic features of social media users' sentiment over popular issues in China. We selected eight popular incidents in China for case studies, and over 200,000 Sina Webos were analyzed. We found that there were three phases of the group sentiment. For each stage, official accounts and grassroots accounts are significantly different. Groups with highly polarized sentiments are detected, posing big challenges for crisis managers. Other managerial insights are discussed.

2 - Optimal News Search and Reporting

Weize Yin, University of Minnesota - Twin Cities, Minneapolis, MN, United States, Yi Zhu

One consumer's belief of an event may frequently fluctuate as she encounters different news reports during her news search on a news medium. In this paper, we investigate how a news medium strategically chooses its news reporting slant to maximize online consumer engagement and, consequently, digital advertising revenue. The consumer not only values more certain updated beliefs due to news search but also prefers news that aligns more with her inherent attitude. By establishing a micro foundation for the consumer's sequential news search, we model a Bayesian consumer's belief update based on Brownian motion. We explicitly characterize the expected duration of the consumer's news search, allowing us to examine the news medium's optimal slant. We find that a consumer stops her news search if her belief becomes sufficiently certain. A sufficiently strong attitude tends to prolong a consumer's news search time, even if the medium's slant opposes this attitude. Counterintuitively, we find that the news medium tends to be more slanted if the consumer values information learned from news search more. This is because the medium can utilize news slant to obfuscate the consumer's information learning, thereby retaining the consumer for a longer duration.

3 - Exploring the Dynamics of Hashtag Use and Comment Toxicity on Social Media

Anna Gao, University of Pennsylvania, Philadelphia, PA, United States

The advent of social media has revolutionized how information is disseminated and discussed globally. This project explores the role of informational cues, such as hashtags used by content creators, in fostering safe online communities, with a particular focus on comment toxicity in social media discussions. Using Twitter's API, we collected tweets and engagement metrics from a week-long sample, and used the Trump Trial as an initial case study. We aim to understand what drives toxic comments during ongoing, high-profile events. In particular, we apply zero-shot learning to classify tweets into categories of support, opposition, or neutrality towards Trump and discover that tweets opposing Trump receive higher engagement and more negative sentiment. Additionally, we assess hashtag relevancy using a pre-trained language model and find that higher relevancy scores correlate with increased tweet impressions, while greater variance in relevancy scores reduces engagement. Our preliminary analysis reveals key insights into hashtag relevancy and comment toxicity. There is a negative correlation between average relevancy and the proportion of toxic comments, indicating that the more relevant the hashtags are to the tweet's content, the smaller the proportion of toxic comments. The findings suggest that strategic use of relevant hashtags can mitigate toxic interactions, fostering healthier online discussion. By understanding and leveraging these dynamics, content creators can make smarter decisions to create a safer digital world.

4 - Tourism Forecasting with Multimodal Social Media Data: An Explainable Machine Learning Framework

Xin Li, University of Science and Technology Beijing, Beijing, China, People's Republic of, Xiangbin Yan

The rapid rise of social media platforms has provided a rich source of multimodal data, containing valuable insights that influence consumer decision-making. While existing research predominantly examines the impact of online reviews on tourism decisions, there is limited exploration into the joint analysis of text and image modalities, resulting in a lack of development of a comprehensive and interpretable analytical framework. This study proposes a novel explainable machine learning framework comprising aspect-based sentiment analysis for textual data, residual neural networks for image features, and long short-term memory models with attention mechanisms for accurate forecasting. Leveraging multimodal data from social media platforms including Ctrip, this research extracts diverse features from online reviews and images, examining how these features influence the forecasting of volumes at Jiuzhaigou tourist attraction. Findings reveal that the incorporation of textual and visual modalities enhances the accuracy of tourism demand forecasting, with varying degrees of improvement in forecasting efficacy attributable to different modalities of information. This research contributes to advancing the understanding of multimodal data analytics and underscores the importance of integrating text and image data for achieving smarter decisions in destination management.

TA63

Regency - 601

Machine Learning in IS: Biases and Model Improvement

Invited Session

Information Systems

Chair: Wei (Weiguang) Wang, Johns Hopkins University, Rochester, NY, United States

1 - When Systemic Biases Taint Algorithms: A Path to More Equitable Access in Healthcare

Ozgur Aksoy, The University of Texas at Dallas, Richardson, TX, United States, Mehmet Ayvaci, Asunur Cezar, Srinivasan Raghunathan

While predictive algorithms hold promise for improving resource allocation decisions in socially critical contexts, they often exhibit "algorithmic bias," inadvertently favoring the socially advantaged. Current bias mitigation strategies typically focus solely on correcting predictions, neglecting downstream effects on subsequent decisions and the underlying societal disparities perpetuating this bias. We investigate the effects of algorithmic bias stemming from societal disparities within an integrated prediction-decision framework. Specifically, we assess how explicitly incorporating social factors into prediction-decision pairs influences outcomes, particularly for a healthcare payer's decision to allocate care management programs to beneficiaries at risk of high costs. We first theoretically model the payer's allocation decision and compare outcomes when adjusting predictions using social factors versus not. While adjustments can mitigate the negative impact of algorithmic bias during the prediction stage, reducing bias does not necessarily decrease disparities in decision outcomes, especially when variations in intervention effectiveness exist between advantaged and disadvantaged. Moreover, tight budget constraints on intervention can diminish the value of adjustments for outcome disparity reduction. Using Medicare insurance payments and CDC Social Vulnerability Index data, we develop a predictive algorithm, quantify algorithmic bias, and propose an adjustment method sensitive to societal disparities. Our results demonstrate this method enhances fairness (i.e., 27.03% increase in the number of disadvantaged beneficiaries among intervened) while simultaneously improving decision efficiency (i.e., replacing 8.20% of intervened beneficiaries with others having 18.65% higher future healthcare costs). Although we focus on health insurance, our approach could apply to other socially and economically significant contexts within and beyond healthcare.

2 - Enhancing Knowledge Retention in Knowledge Graph

Chaofan Zhai, University of Minnesota, Minneapolis, MN, United States, Yicheng Song, Ravi Bapna

Our research investigates a deep-learning framework to improve knowledge retention on online learning platforms. We address challenges in predicting knowledge retention, engagement, and optimizing long-term learning outcomes using a combination of knowledge graphs and transformer models. Our approach integrates static and dynamic knowledge graphs with embeddings derived from user interactions and predefined semantics. Additionally, we employ reinforcement learning to recommend learning paths that balance engagement and performance. Empirical results from MaiMemo.com, a major English learning platform in China, show our KG_Trans model achieves superior performance with an accuracy of 0.792 and an F1 score of 0.858. Future work will evaluate the performance of survival models and reinforcement learning in real-world settings. This research contributes to personalized and effective online learning, potentially applicable to various educational contexts.

3 - Key Verbatim Extraction from Clinical Notes: A Hierarchical Multimodal Cross-Attention Approach

Feiyu E, Robert H. Smith School of Business, University of Maryland, College Park, MD, United States, Hanwen Shi, Jin Zhang, Kunpeng Zhang

Clinical notes are pivotal for physicians to accurately assess patient conditions, particularly in oncology where records are extensive. Efficient information extraction from these notes is crucial for effective treatment. Quickly understanding clinical notes remains challenging due to their specialized content and extensive length. Current methods do not fully utilize relational information among sentences or across different notes, and the exploration of relationships among health-related entities is insufficient. We introduce a hierarchical multimodal cross-attention model that utilizes a cross-attention mechanism to integrate textual knowledge with patient network knowledge. It aims to synthesize information across various data levels, including word, sentence, note, and patient level, to efficiently highlight key sentences in clinical notes. Initial testing shows the model achieving a higher accuracy compared to baseline models, indicating effective learning but highlighting a need for better generalization. Enhancements will include integrating multimodal data, refining network knowledge, implementing regularization techniques, and optimizing learning parameters to improve model generalizability. This study offers a novel approach to clinical text analysis by combining multiple data levels and utilizing inter-patient relationships, potentially improving healthcare delivery and decision-making efficiency.

4 - The Risk of Inferring Data Insights from Post Hoc Explanations of Machine Learning Models

Ronilo Ragodos, University Of Iowa, Iowa City, IA, United States, Tong Wang, Jeffrey Hu, Lu Feng

The integration of machine learning (ML) methods in business research has seen a notable rise, prompted by the necessity to analyze complex datasets. While ML models are adept at making predictions, business researchers often seek an understanding of the underlying data. As post hoc explainers, such as SHAP and LIME, have emerged as popular tools to explain ML models, it has become increasingly common in business research for authors to use explainers to make inferences about their datasets rather than their models. This paper critically examines this trend. Specifically, we investigate whether post hoc explanations contain three types of information about the marginal effects of features in data that researchers seek, based on recent literature. We formalize the three types of information and refer to explanations that can capture them as data-aligned. We identify factors within the explanation pipeline that are associated with explanations that are not data-aligned. We find ideal conditions where LIME explanations are data-aligned, and we show theoretically that SHAP is inherently unsuited for recovering marginal effects. We also discuss the issue of explanation inconsistency, which is due to varying choices of models, explainers, and their configurations. Furthermore, we propose simple mitigation strategies for low data-alignment and inconsistency that we encourage researchers to use as a foundation for experimentation. We validate our mitigation strategies in an econometric context and demonstrate the resulting improvement in explanation quality. Despite these efforts, we nevertheless conclude that researchers should only use post hoc explainers to infer data insights with great caution.

TA64

Regency - 602

Unlocking the Digital Discourse: From Data to Insights

Invited Session

Social Media Analytics

Chair: Priyanga Gunarathne, University of Pittsburgh, Pittsburgh, PA, United States

1 - Social Interaction Network Driving Purchase Behavior

Yoonjin Choi, Georgia Institute of Technology, Atlanta, GA, United States, Mingfeng Lin

E-commerce is increasingly social, with many different channels through which influencers and consumers may interact with each other, either before or after a purchase. Existing research, due to a lack of data or methodological constraints, often cannot fully account for the various ways through which online purchases take place. We investigate the extent to which purchases occur after online social activities and how we can leverage the relationships among different types of online social activities to improve the identification of influencers and the prediction of sales. We leverage a comprehensive dataset from an e-commerce platform where consumers can either purchase products directly or follow, like, or comment on the postings of influencers; the influencers may also directly interact with each other. We apply the Relational Graph Convolutional Networks method (R-GCN), which considers the multiplex nature of the networks, to address these tasks. Not only do these models improve the performance of prediction models based on single-layer network data, but they also yield important practical insights.

2 - From Self to Other: Customer Reactions to Firm DEI Promotions on Social Media

Chen Jing, Boston University, Boston, MA, United States, Priyanga Gunarathne, Shuba Srinivasan, Abraham Seidmann

Our paper investigates customer reactions to firm promotions of Diversity, Equity, and Inclusion (DEI) on social media. We assess how these initiatives affect customer communications, including complaints, and how perceptions vary across different demographic groups. Utilizing a dataset comprising public tweets from 15 major U.S. firms across various industries, collected from May 2020 to March 2023, we apply advanced Natural Language Processing (NLP) methods and face recognition technologies to classify DEI content and link it to specific customer identities. Through fixed-effects Poisson regression and the instrumental variables approach, our analysis reveals that DEI promotions are generally not well received by consumers, potentially exacerbating customer complaints even when paired with improvements in a firm's functional services. Our results also identify different impacts of DEI promotions based on their alignment with customer identities: promotions that coincide with customers' identities positively influence perceptions and reduce customer complaints. Additionally, we find that non-corresponding DEI promotions are the main source of adverse reactions, highlighting the critical need for alignment between DEI efforts and the distinct identities of customer segments. These insights underscore the nuanced challenges firms face in implementing DEI strategies that resonate positively across their diverse customer bases.

3 - Gender, Knowledge, and Engagement in Digital Communities of Practice: Evidence from X

Priyanga Gunarathne, University of Pittsburgh, Pittsburgh, PA, United States, Ruba Aljafari, Sezgin Ayabakan, Samer Khader, İbrahim Kulaç

Physicians are increasingly harnessing social media for professional networking, community interaction, and outreach. Unlike offline interactions within organizational hierarchies, where peers recognize each other's competencies and backgrounds, individual characteristics perceived on social media may shape online interactions, influencing career trajectories and peer recognition. Specifically, it remains unclear whether gender emerges as an individual characteristic influencing community engagement in digital physician communities on social media, as such online interactions involve nuances related to engagement with professional knowledge. Using the pathology community on X (formerly Twitter) as our research context, we exploit a novel dataset of professional conversations among pathologists on X and examine the role of gender in shaping peer engagement with professional knowledge. We leverage 2,467 patient case-based tweets posted by pathologists on X over a period of six months and employ a holistic set of tweet- and pathologist-specific controls and a matched sample approach as our empirical strategy. Our findings provide quantitative evidence of more favorable community engagement outcomes (likes, retweets, and replies) for professional knowledge created by female pathologists on X, compared to similar content created by similar male pathologists. To theoretically explore the identified gender differences, we surveyed pathologists on X. Our survey findings reveal that female pathologists are perceived as more communal, cooperative, and helpful, whereas male pathologists are viewed as more dominant and competitive. Our findings provide key insights for various stakeholders, including physicians, clinicians, hospital administrators, and platform owners who aim to foster, harness, and potentially to regulate communities of practice on digital platforms.

TA65

Regency - 603

Community-Engaged and Equity-Forward Operations Research and Analytics II

Invited Session

Public Sector OR

Chair: Michael Johnson, University of Massachusetts Boston, Boston, MA, 02125, United States

1 - Simulation Optimization of Queueing Models for Capacity Planning in Homelessness Care Systems

Dashi Singham, Naval Postgraduate School, Monterey, CA, United States, Graham Burgess

A leading approach to resolving homelessness in highly populated areas is to increase the amount of housing and shelter available. Alameda County, located in the San Francisco Bay Area, has a focus on racial equity in its recent homeless response system redesign. By investing in multiple classes of housing to address different types of needs, Alameda County aims to address the unique needs of racial minorities who are overrepresented in the homeless population. We build a simulation models to assess the effects of investment levels in critical pathways designed to reduce the unsheltered population. Simulation optimization can be used to propose investment plans to best reach the housing goals of Alameda County's racial equity plan.

2 - Navigating Health Outcomes for Transgender Individuals: The Influence of State Legislation

Yuehwern Yih, Purdue University, West Lafayette, IN, United States, Min Kyung Lee

Addressing health disparities among transgender and gender-diverse (TGD) populations is a key public health concern. This study evaluates the dynamic causal relationships between state-level legislation and health outcomes for TGD individuals in the United States, employing a marginal structural model with inverse probability weighting to account for time-varying confounding, such as sociodemographic factors. We utilized data from the Behavioral Risk Factor Surveillance System (BRFSS) and the Human Rights Campaign's State Equality Index (HRC SEI), which provide comprehensive insights into health behaviors and legislative climates, respectively. Our findings indicate that discriminatory policies significantly exacerbate health disparities, leading to poorer mental and physical health outcomes among TGD populations. Notably, the interaction between legislation and transgender status significantly affects mental health, with unsupportive legislative environments intensifying negative impacts on both physical and general health for transgender respondents, whereas these legislative factors do not significantly impact the health outcomes of cisgender respondents. This study underscores the importance of supportive legislation in mitigating health disparities faced by TGD individuals and highlights the need for policies that promote health equity.

3 - Revisiting equity in urban operations management 50 years later: What do city planners have to say?

Mark Brennan, Rutgers University, Camden, NJ, United States

The persistent inequities in American cities—long recognized and lived by those disadvantaged by urban systems—have been vaulted into the broader public consciousness over the past decade. Lessening entrenched urban inequalities is now at the top of the national policy agenda, suggesting a need and opportunity for more urban operations management. On what issues and how might this work occur? Operations management was deeply intertwined with urban planning from the 1950s through the 1970s, at which point the fields diverged. To build a case for what perspectives a modern urban operations management agenda might employ to address inequity, I synthesize historical and current planning thinking with the debates among reflective urban operations scholars during 1950s-1970s. This urban operations agenda should be empirical, equity-oriented, and community-focused in order to resonate with city planners and the residents they serve. Modern operations scholars can look to new thinking in planning on race and class in order to address urban operations that especially disadvantage some city residents and overly advantage others. In reengaging with planners to tackle modern urban policy problems, operations analysts have a chance to contribute practical clarity on how cities work and can be made more livable for all.

TA66

Regency - 604

Enhancing Supply Chain and Operations Efficiency: AI-driven Inventory Management

Invited Session

Artificial Intelligence

Chair: Zheng Zeng, Rutgers University, East Brunswick, NJ, United States

1 - A Queueing System with a Parallel Auxiliary Service

Peng Li, Rutgers Business School, Newark, NJ, United States, Chunliu Zhou, Lian Qi

This paper examines the maintenance and repair industry's practice of servicing high-value components of long-life-cycle machines. During machine servicing at primary stations, these components are removed and serviced at auxiliary stations, then reinstalled. To optimize machine utilization and minimize downtime, an inventory of fully serviced parts is maintained, allowing immediate swapping with depleted parts. We model this system as a queueing problem with primary and parallel auxiliary processes, focusing on optimal investment in primary and auxiliary servers and the inventory of backup parts, considering uncertainties in customer arrivals and service times. Our research uniquely optimizes investment in queueing systems with auxiliary servers and backup stocks. We establish the conditional probability of serviced part availability when customer demand exceeds primary server capacity and define stability conditions for the system. An approximation system is developed to provide near-optimal solutions with less than 5% numerical error, offering valuable managerial insights. This study contributes significantly to improving operational efficiency, service delivery, and customer satisfaction in maintenance and repair operations by optimizing resource allocation and management.

2 - Optimizing Production Capacity: Insights for Manufacturing Companies

Zheng Zeng, Rutgers University, Edison, NJ, United States, Xiaowei Xu

In the ever-changing market landscape, companies face the challenge of balancing customer demands for both immediate delivery and customizable products. In addressing the operational optimization challenges, this study adopts a robust methodology combining two-stage stochastic programming with the Newsvendor approximation.

The two-stage stochastic programming framework accommodates the inherent uncertainty in demand forecasting and production outcomes. In the first stage, decisions are made based on available information, such as prebuilt SKU inventory levels and production capacity. Assortment planning for top-selling SKUs is optimized to balance production constraints and demand variability, aiming to minimize costs while meeting customer requirements. In the second stage, which accounts for uncertainty, decisions are adjusted based on actual demand realizations. Here, the Newsvendor approximation is utilized to model the uncertain demand for immediate delivery and the potential for substitutions. The Newsvendor approximation helps to mitigate the risk of stockouts while minimizing excess inventory holding costs.

In addition, this study involved the practical application of the proposed approach using real-time data to facilitate a company's transition from a make-to-order to a make-to-stock model. The application of the methodology on real-time data allowed for a pilot implementation, providing the company with actionable insights and tangible results. By leveraging actual demand data, production capacity, and inventory levels, the study enabled the company to assess the feasibility and effectiveness of transitioning to a make-to-stock approach.

3 - Periodic Review Inventory Models with All-or-nothing Yield - A Transformer Framework

Zhe FU, City University of Hong Kong, Hong Kong, Hong Kong, Frank Chen

Training a deep reinforcement learning (DRL) model takes much time, including processes like storing computational graphs and computing gradients during forward and backward propagation. For inventory management problems, the states in Markov Decision Processes (MDP) typically include variables such as inventory levels, order quantities, and more. In many studies, the DRL agents utilize such states as input to take optimal actions like inventory adjustment strategies or orderings to maximize cumulative rewards. This paper proposes a novel approach to establish a pass-through connection between rewards and stochastic factors such as demands and random yields using Transformers, which is a class of deep neural networks (DNNs) with significant advantages in handling long-term dependencies and parallel processing. By integrating the MDP environment and DRL agents into Transformers’ architecture, our framework can process states that only contain random variables and directly output cumulative rewards. We utilize this framework with deep reinforcement learning to solve a single-item periodic review inventory problem with random demand, an all-or-nothing random yield pattern, and lead times. Its optimal structure is unknown, and the exact solution suffers from the curse of dimensionality. Our framework can handle these problems more efficiently and significantly improves computational speed compared to the general DRL models with inventory-related states. Moreover, this framework supports Transfer Learning, where the teacher's policies such as heuristic methods for the inventory problem are applied to train a model. Other applications of the proposed framework will also be discussed.

4 - SCGNN: Forecasting Food Prices with Constructed Supply Chain Graph and adopted Graph Neural Network Model

Fuqin Zhou, New Jersey Institute of Technology, Newark, NJ, United States, Jiaxing Zhang, Aichih Chang

The fluctuating prices of fresh food have created challenges for consumers and stakeholders after the pandemic. Existing methods treat the data as time sequences, which fail to measure the relationship between the markets and the impact of the supply chain network. To address this issue, we need a predictive model that can forecast prices on a large scale, taking into account factors such as spatial origins, transportation networks, terminal markets, and food categories. In our study, we gathered a spatial-temporal dataset of fresh food prices in US markets. We adopted a spatial-temporal Graph Neural Network (ST-GNN) framework to create a comprehensive solution for various forecasting tasks, including a gated message-passing mechanism to capture the inherent interdependencies in food prices. Our solutions aim to contribute to stability and resilience in the agricultural food market ecosystem.

TA67

Regency - 605

Value Creation in Online Platforms

Contributed Session

Chair: Prabin Panigrahi, Indian Institute of Management Indore, Indore, India

1 - Rebate policies in the presence of social learning and strategic consumers

Bingyue Hu, Tianjin University, Tianjin, China, People's Republic of, Yuan Chen, Yinghua Shen

With the prevalence of online shopping, firms have recognized the influential role of online reviews in shaping consumers' beliefs about product quality. To capitalize on this, firms have employed rebate strategies incentivizing consumers to write reviews and even positive reviews. Note however, with the increasing rebate amount of firms, consumers may anticipate potential benefits of using more generated reviews to reduce their uncertainty of product quality, leading more consumers strategically delay their purchasing decisions. This dilemma makes it unclear for firms to determine right rebate policies. This paper employs a game theoretical model to investigate the interaction between firms and strategic consumers within the framework of Social Learning (SL), where consumers update their beliefs about product quality through Bayesian rule after observing reviews from previous purchasers. We develop a two-period stylized model to analyze sellers' pricing and rebate strategies (conditioned on writing reviews or positive reviews), considering the strategic decisions of consumers on when to purchase and whether to write reviews. Our research reveals that rebate policies are beneficial since SL is advantageous for both consumers and firms when there is no information distortion or only slight information distortion. This environment facilitates well-informed purchase decisions for consumers, while firms gain a better understanding of consumers' product quality beliefs. However, we also examine how the extent of information distortion affects firm profit and consumer surplus. Our findings indicate that information distortion invariably benefits firm profit and detrimentally impacts consumer welfare; but this happens only when penalty of review manipulation is low.

2 - Fast solution to the fair ranking problem using the Sinkhorn algorithm

Yuki Uehara, University of Tsukuba, Tsukuba-shi, Japan, Shunnosuke Ikeda, Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano

Ranking algorithms play an important role in e-commerce recommendation systems. Traditional rankings focused only on buyer satisfaction, but recently it has become increasingly important to consider the satisfaction of item providers. Rankings that aim to fairly increase the satisfaction level of item providers are called fair rankings. The fair ranking problem for leveling the expected impact on each item is formulated as a large-scale constrained nonlinear optimization problem, but it is time-consuming to solve this problem using commercial optimization solvers. To efficiently solve the fair ranking problem, we propose a gradient ascent method that is highly compatible with parallel processing. To this end, we reformulate the problem as an unconstrained problem involving a projection function, which is implemented by the Sinkhorn algorithm in our method. The Sinkhorn algorithm is suitable for parallel processing and provides gradient information for the reformulated problem. Numerical results demonstrate that our method is significantly faster than applying a commercial optimization solver to the existing formulation, without compromising solution accuracy.

3 - Fake Reviews are out of Depth: Fake Review Detection using Semantic Networks.

Prabin Kumar Panigrahi, Indian Institute of Management Indore, Indore, India, Abhishek Kumar Jha

In the digital age, online consumer reviews (OCRs) significantly influence purchasing decisions, making platforms vulnerable to fake reviews. This study aims to improve fake review detection by analyzing the semantic differences between fake and real reviews using network text analysis. The research first identifies these differences in an uncontrolled eCommerce environment. Reviews from Amazon were pre-processed and analyzed, with each review represented as a network node connected by edges based on cosine similarity. Network properties such as density, connected components, and clustering coefficients were evaluated to distinguish fake from real reviews. Incorporating these network-derived parameters into machine-learning models significantly improved detection accuracy. The findings reveal that fake reviews exhibit higher semantic similarity within product categories, enhancing model accuracy for identifying fake reviews, among others. This novel approach provides both theoretical insights about the characteristics and practical tools for better managing the authenticity of online reviews, contributing to a more trustworthy eCommerce environment. The underlying findings also indicated in the conversation that the reviews, when real, are more specific, leading to a lower similarity.

In contrast, the fake review only repeats the conversation about the highlighted product features without more variation. This is a theoretical finding that we believe is true for another form of misinformation across different networks. Limitations include the use of cosine similarity and testing on local networks, suggesting future research could explore alternative measures and broader network scopes.

Keywords: Fake Review, Semantic Network, Network Analytics, Machine learning

TA68

Regency - 606

Innovations in AI and Digital Engagement

Contributed Session

Chair: Manni Gao, Harbin Institute of Technology, Harbin, N/A

1 - Discovering Convergence of AI Technology and their Spreads: Network Perspectives

Won Sang Lee, Gangneung Wonju National University, Gangneung, Korea, Republic of

Currently, the convergences with AI technology accelerate diverse interactions and continuous advancements in many areas, thereby attracting great interests of both researchers and practitioners. Such technological convergences on AI and their spreads unprecedently occur in the complicated phases. The necessity of comprehensively investigation is increasing. This paper focuses on what technologies have been composing AI convergences and how those spread into diverse domains based on the entire triadic patents. The network dynamics and diffusion model are jointly deployed to track such spreads of AI with emphasis on the impact of the technological convergences of AI. In particular, it is further concerned whether these convergences benefit from the inflow of external technology or affect other technologies during the spread of AI convergences. The proposed methodology is expected to discover the promising AI convergences. Findings of this paper could, then, lead us to acknowledge the intrinsic nature of AI convergences and provide the policy implication for their utilization and dissemination.

2 - Does content matter? The influence of fine-grained, textual aspects on review helpfulness.

Alexander Kupfer, University of Innsbruck, Innsbruck, Austria, Stefanie Erlebach, Kilian Züllig, Leonie Embacher, Steffen Zimmermann

Even though helpfulness votes in online review systems predominantly aim to facilitate the consumers' identification of the most helpful opinions on a product (Mudambi and Schuff, 2010), the consequential data availability of online consumer reviews that are voted as helpful has also attracted many scholars in Information Systems research. Many studies use these votes as dependent variables to examine the relevance of various factors that "make" a review helpful (see Rietsche et al., 2019, for an overview). The use of these votes, however, restricts the analysis' effectiveness in terms of granularity, and one consequently observes rather "broad" metrics like, for instance, textual length or overall sentiment to be correlated with helpfulness votes. Even though one would expect that the actual content of a review influences its helpfulness, there is no evidence on how fine-grained, textual aspects affect review helpfulness. To address this lack of knowledge, we used a real-world review dataset, classified the respective reviews textual content along four dimensions (product-related, customer-related, seller-related, and purchase condition-related) including the dimensions' sentiment, and applied a discrete choice experiment to extract fine-grained helpfulness preferences. In this context, we transfer the feature of discrete choice experiments to extract preferences for specific product characteristics to identify preferences for specific textual characteristics in reviews. Our preliminary findings indicate that – as expected – particularly product-related aspects (positive and negative) are helpful. However, we also document that, e.g., purchase condition-related aspects are not helpful. Having fine-grained information about review helpfulness is essential for academics and businesses alike.

3 - Loyal Costumer, Brand Defender or Betrayer? Evidence From Sequential Scandals in Multi-Brand Apparel Firms

Yi Zeng, City University of Hong Kong, Hong Kong, China, People's Republic of

Brand scandals have always been an interesting and commercially valuable research topic. After a scandal, consumers change their attitudes towards the brand and their purchasing behavior, which in turn affects brand equity and profitability. Previous research has shown that consumers with different characteristics will react differently to the same scandal, and understanding the differences and underlying mechanisms can help brands apply different strategies to different groups when it comes to remediation.

By using the DID model, we analyzed the types of consumers and their purchasing data before and after two consecutive scandals in a Chinese multi-brand clothing company and conducted a robustness test. Preliminary results suggest that loyal consumers are more likely to reduce consumption and disassociate from brands when faced with value-related scandals. We also find that this effect is influenced by consumers' attribution: when the scandal of sub-brand is attributed to the parent brand, the parent brand not only suffers negative spillover effect, but also the betrayal extent of its loyal customers is greater; When the sub-brand scandal is attributed to itself, the parent brand's loyal customers appear to defend and the sub-brand's loyal customers appear to betray. In addition, combined with the text analysis of the related posts on Weibo, we also found the influence of social media in the fermentation of the scandal.

4 - Specialization or Diversification? Creators' Strategies and User Engagement in User-Generated Content (UGC) Platforms

Ziwei Ye, Old Dominion University, Norfolk, VA, United States, Kedong Chen, Lan Cao

Recent advancements in social media have revolutionized content creation and distribution. User Generated Content (UGC), consisting of original content created and shared by internet users, is reshaping the realms of entertainment, communication, and information. To maintain a competitive advantage, UGC creators may strategically adopt specialization or diversification in their content creation. Specialization refers to a focus on "only one type of task" (e.g., a specific niche, topic, or content category within the platform), whereas diversification engages in multiple tasks, seeking to capitalize on the competitive advantages of being a "one-stop shop" by offering a diverse range of products and services. While prior studies have examined the impact of specialization degrees on creators' follower growth, the existing literature offers limited evidence on how such strategies either enhance or diminish user engagement. This study leverages a substantial panel dataset from a popular short video platform to examine how creators' specialization versus diversification strategies affect user acquisition and retention. Employing fixed effects models, we find that the degree of creators' specialization is positively associated with their user acquisition rate while negatively associated with their user retention rate. Furthermore, the effect of creators' specialization strategy on user engagement is moderated by platform's algorithmic intervention. These findings underscore the trade-off between specialized and diversified content strategies in shaping user engagement metrics. By understanding these dynamics, platforms can better guide creators in selecting strategies that optimally balance user acquisition and retention, thereby maximizing overall user engagement.

5 - Empowering Innovation: An Empirical Study on the Impact of Generative AIon Online Freelancers' Performance

Manni Gao, Harbin institute of Technology, Harbin, China, People's Republic of, Jianwei Liu, Qiang Ye

Generative AI is revolutionizing content generation, particularly redefining success for freelancers engaged in image creation. This study examines the impact of generative AI adoption on freelancers' performance and dissects the underlying mechanisms. Leveraging extensive data from freelancers specializing in image creation on a premier global gig economy platform, we adopt a difference-in-differences methodology for empirical analysis and use deep learning techniques to assess specific image attributes, such as color and content diversity. We have three main findings. First, adopting generative AI improves freelancers' performance in color and content diversity of their generated images while also expanding their order volumes and customer reach. Second, the mechanism centers on the limited lifting effect of generative AI on human-generated images, which notably increases order volumes and content diversity without altering the fundamental styles, except for those with relatively singular painting styles. Third, content diversity empowered by generative AI is a strong driver of growth in traditional manual services, but the reason behind overall services growth is relatively mixed. Our research contributes to two critical areas: (1) the examination of technological evolution's impact on online labor markets, an area underexplored in existing literature, and (2) the implications of generative AI on content creation, where this emerging technology is reshaping creative paradigms. This research provides insights for administrators and policymakers on adapting to the emerging technology and making informed decisions.

TA69

Regency - 607

E-commerce and Supply Chain Management

Invited Session

eBusiness

Chair: Emily Griffin, Babson College, Wellesley, MA, United States

1 - Online Supplier Selection

Thomas Cassidey, Cal State, East Bay, Hayward, CA, United States, Emily Griffin, Burcu Keskin

We address the problem of an e-retailer that connects drop shipping suppliers to customers. The retailer must balance the potential tradeoffs between cost and delivery performance on each order. We develop and compare several algorithms to manage this tradeoff and present empirical results.

2 - Bargaining Over Data and Analytics: Sellers, Buyers and Consultants

Jyotishka Ray, University of Dayton, Dayton, OH, United States

The explosive growth of online commerce has generated large quantities of data that can be used by firms to improve decision making. Some of the data collected can be directly used by firms, for example, in an advertising campaign to target certain users. In other cases, these data sets can be further analyzed using sophisticated analytic methods to substantially increase their value.
We examine the selling of data where the value to the data buyer can be enhanced with the use of analytic services obtained from a consultant. Because our context emphasizes the exclusive selling of proprietary and unique data (rather than general purpose data that can be sold to many buyers), announcing a fixed price for the data is not a viable option for the seller. Thus we use a Nash bargaining framework where the negotiations always involve the data seller and buyer and could sometimes involve a consultant. Data sellers can choose to sell the data alone or sell a bundle that includes the data and complementary analytic services. Data buyers can choose to simultaneously negotiate the price of data and the price of analytic services. Alternatively the data buyer could perform these negotiations in two separate steps. Our analysis shows that the contribution of the consultant's analytic services is critical to both the seller's decision to bundle and the buyer's choice over how best to structure the negotiations. Data sellers have a natural advantage over external consultants when they choose to bundle data with complementary analytic services.

3 - How many software versions to support: The role of software support experience & cost

Jayarajan Samuel, The University of Texas at Arlington, Arlington, TX, United States, Alper Nakkas

Commercial off the shelf (COTS) software form the backbone of any e-business platform. The product support of these COTS software products is a profitable and cost sensitive business. We investigate, though a unique dataset, the effect of software versions on user support experience and costs. We find that keeping customer software deployments in homogenous software versions is challenging but has high rewards. We further draw causal insights on the effects of divergent software release deployments on cost.

4 - Return Freight Insurance Strategy in Competitive E-Commerce Operations

Shan LI, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

In response to consumers' pursuit of free return shipping, Chinese e-commerce platforms provide free return insurance to consumers. Meanwhile, with the increasing encroachment of platform-operated e-tailer in the market and its emergence as competitors to the manufacturer in the downstream supply chain, a crucial question arises: when should firms provide free return freight insurance (RFI) to consumers in this new environment? Our research reveals that offering RFI leads to an increase in demand for the firm and a decrease in demand for rival firms, which we term the *market expansion effect* caused by RFI. We also find that manufacturers are more motivated to offer RFI compared to platform-operated e-tailers. Additionally, both of these two firms tend to set a higher retail price when they provide RFI to consumers, regardless of the insurance premium's value. This phenomenon, which we refer to as the *cost enhancement effect*. Furthermore, we find that when the insurance premium is low, both of these two firms would provide return freight insurance. However, as the premium increases, neither party would offer return freight insurance to consumers. Interestingly, in cases where the insurance premium is high and the rival firm does not offer RFI, both firms surprisingly benefit from the decision to provide RFI. This is due to the significant *market expansion effect* experienced by the firm offering RFI under this circumstance, resulting in benefits for the e-tailer even with a sufficiently large premium. We term this phenomenon the *mutual exclusive effect*.

TA70

Regency - 701

Advances in Sustainable Energy Systems Operations

Invited Session

ENRE: Electricity

Chair: Kai Pan, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Co-Chair: Bin Tian, N/A

1 - Operationalizing Smart Charging; a Robust Optimization Interpretation of Flexibility Vs Uncertainty

Linda Punt, Rotterdam School of Management, Rotterdam, Netherlands, Yashar Ghiassi, Gar Goei Loke

The electrification of transportation and increasing electric vehicle (EV) adoption poses challenges for local power grid management. Our research addresses these challenges by (i) proposing a decision-support model in the form of a robust optimization framework for implementing different smart charging regimes, and (ii) analyzing the interplay between uncertainty and flexibility in smart charging. Specifically, we propose an aggregate model for EV arrivals and departures and prove that individual EV charging schedules can be recovered from their solutions. We also propose a novel uncertainty set that incorporates length of stay (LOS) to relate EV arrivals and departures. Through our model, we offer insights into various flexibility options, revealing that smart charging, even when LOS is exogenously determined, can prove more cost-effective than no smart charging for sufficiently high rates of EV adoption. We also emphasize the importance of optimal external grid capacity and battery investments, showcasing their complementary relationship with EV adoption rates under smart charging, and note that increased investments in these complementary flexibility instruments enhance the cost-effectiveness of implementing smart charging for operators. Additionally, our analysis identifies batteries as the most cost-effective solution to mitigate EV departure uncertainty.

2 - Charging An Electric Vehicle on Fast-Lane and Station: Convenience, Accessibility and Availability

Yanlu Zhao, Duham University Business School, Durham, United Kingdom, Kai Pan

The rapid increase in electric vehicle (EV) use has profoundly transformed urban transportation and placed significant demands on the energy system, primarily due to the increased need for battery charging. Stationary charging limits vehicle availability, particularly for heavy goods vehicles (HGVs) that require substantial energy for long-haul trips with heavy loads. This requirement adds to the vehicle weight, perpetuating a problematic cycle. In response, some pioneering countries like Sweden and innovative companies are exploring road-charging technologies that charge EVs while they are moving, thus reducing downtime for charging and bringing this strategy into the spotlight. Our research focuses on this innovative charging approach by examining a new problem that integrates both road and stationary charging methods. We utilize mixed-integer programming to address this issue and broaden our investigation to account for uncertainties in charging station availability and variable energy prices. We adopt a distributionally robust approach to develop broader strategies for more practical contexts. The effectiveness of our solutions is validated using real data from China, demonstrating the significant benefits from mixed charging strategies. We also obtained more managerial insights that guides the charging station design and energy budget management.

3 - A Polyhedral Study on Unit Commitment with A Single Type of Binary Variables

Bin Tian, The Hong Kong Polytechnic University, KOWLOON, Hong Kong, Kai Pan, Chung-Lun Li

Efficient power production scheduling is a crucial concern for power system operators aiming to minimize operational costs. Previous mixed-integer linear programming formulations for unit commitment (UC) problems have primarily used two or three types of binary variables. The investigation of strong formulations with a single type of binary variables has been limited, as it is believed to be challenging to derive strong valid inequalities using fewer binary variables, and the reduction of number of binary variables is often accompanied by a compromise in tightness. To address these issues, this paper considers a formulation for unit commitment using a single type of binary variables and develops strong valid inequality families to enhance the tightness of the formulation. Conditions under which these strong valid inequalities serve as facet-defining inequalities for the single-generator UC polytope are provided. For those large-size valid inequality families, the existence of efficient separation algorithms for determining the most violated inequalities is also discussed. The effectiveness of the proposed single-binary formulation and strong valid inequalities is demonstrated through computational experiments on network-constrained UC problems. The results indicate that the strong valid inequalities presented in this paper are effective in solving UC problems and can also be applied to UC formulations that contain more than one type of binary variables.

4 - Decarbonizing Mining Operations: Co-Optimization Strategies for Clean Energy Integration and Operations Planning

Vivienne Liu, National Renewable Energy Laboratory , Golden, CO, United States, Thomas Jeffery, Sourabh Dalvi, Gord Stephen, Devon Sigler

Mining is among the most energy-intensive activities in industry, and a notable remaining challenge in transitioning to fully decarbonized energy systems. This work develops a co-optimization model to plan and operate a fully decarbonized iron ore mining operation, by electrifying diesel-powered mining vehicles to enable them to run on clean electricity resources such as wind, solar, and battery storage. A significant portion of the mining system's electricity consumption stems from mining trucks and processing facility operations, offering substantial demand-side flexibility provided by responsive EV charging and ore processing schedules. Moreover, the stockpiles for mining products provide extra energy storage options and serve as a buffer during extreme events. Therefore, coordinated planning and operation of mining facilities and the power system are vital for cost-effectively achieving decarbonization goals in the mining sector.

In collaboration with Fortescue, we develop a co-optimization model that incorporates mining operation constraints for use with power system production cost model. The results showcase viable strategies to reach clean energy targets and offer insights into facility expansions for both mining operations and renewable energy integration.

5 - Hybrid Quantum Classical Computing for Energy System Management

Lei Fan, University of Houston, Houston, TX, United States, Zhongqi Zhao, Zhu Han

In this talk, we will discuss how to utilize the hybrid quantum classical computing techniques to solve the energy management challenges. We will introduce the hybrid quantum classical Benders' decomposition algorithms and its application in energy management system. We will show the simulation results to verify the our proposed algorithm and the advantages of the quantum computing on the cloud.

TA71

Regency - 702

Integration of Human, Knowledge and Systems for Quality

Invited Session

Data Mining

Chair: Feng Lin, University of Washington, Mountlake Terrace, WA, United States

1 - Trust-aware Human Robot Interaction in Manufacturing Plant

Jundi Liu, Iowa State University, Ames, IA, United States, Mobina Amorollahi, Rindirisia Wangira

Large industrial robots are a cornerstone in contemporary manufacturing, undertaking tasks that demand heavy manual effort or monotonous repetition. Among these, Automated Guided Vehicles (AGVs) have been deployed in numerous manufacturing settings, streamlining the transportation of parts for subsequent processing or operations. Despite their prevalence, these vehicles typically operate in isolated zones, separate from human workers. This segregation stems from two primary concerns: firstly, there's a notable lack of trust among workers towards these mobile robots within safety-critical environments, which hinders the adoption of such systems; secondly, the uncertain nature of human behavior introduces variables that current AGVs struggle to accommodate. Current research often treats human as a data generator and builds data-driven models that lack theoretical understanding of how cognition guides the behaviors. In this study, we utilized Virtual Reality (VR) to create a simulated manufacturing environment to collect data on human behavior while interacting with AGVs. We then analyzed this behavioral data and developed a trust prediction model based on these behaviors to address the inaccuracies often found in self-reported trust measures. This trust prediction model has significant implications for the design of future trust-aware adaptive systems, ultimately enhancing human-robot interaction.

2 - Quality Control in Crowdsourcing based on Random Utility Modeling

Hanming Zheng, City University of Hong Kong, Hong Kong, Hong Kong, Shuai Huang, Li Zeng

Crowdsourcing has emerged as a popular way to conduct complex tasks by combining the contributions of a large group of workers. One challenge in crowdsourcing is quality control, i.e., aggregating the outputs of workers considering worker variations to ensure reliable crowdsourcing results. Some quality control methods have been developed in recent years for certain type of tasks such as binary labelling, but there lacks a generic approach that works for various tasks. This study proposes a novel, simple statistical framework for quality control in crowdsourcing to address the research gap. The proposed method is based on a random utility modelling (RUM) of worker outputs and affecting factors such as worker ability and task difficulty level. Based on the model, subgroups of high-quality workers who will be relied on to generate final result for the given task are further identified via density-based clustering. Although illustrated using binary labelling tasks in this study, this method also applies for other types of crowdsourcing tasks including multi-category labelling, rating, etc., thanks to the generality of RUM. The case study shows that this method is robust across different scenarios and especially it outperforms existing methods when worker assignment is sparse.

3 - Modeling dynamic disease-behavior feedbacks for improved epidemic prediction and response

Hongru Du, Johns Hopkins University, Baltimore, MD, United States, Lauren Gardner, Alison Hill, Sara Loo, Nicholas Papageorge, Bryan Patenaude, Shaun Truelove, Mattew Zahn

Infectious disease transmission and control are deeply intertwined with human behavior. Traditional epidemiological models, while valuable, often overlook the complex decisions individuals make in response to changing disease risks and policies. This study introduces a novel framework that integrates individual decision-making with disease transmission dynamics. By incorporating feedback loops, externalities, and diverse decision-making processes, our model offers a more comprehensive understanding of the interplay between behavior, disease spread, and policy interventions. We demonstrate the capability of this framework by examining how individual decisions influence disease transmission and the subsequent impact of policy choices on health outcomes and work participation in scenarios mirroring COVID-19-like outbreaks. This approach allows us to assess not only epidemiological outcomes but also the economic costs and welfare implications of interventions, particularly for diverse sub-populations with varying health and economic vulnerabilities. Our findings emphasize the crucial role of considering endogenous behavioral responses in policy design. They highlight the potential for more targeted, equitable interventions that effectively balance public health objectives with individual well-being. By integrating human behavior into epidemiological models, we pave the way for more effective and nuanced strategies to combat infectious diseases.

4 - A Rawlsian Mixed Integer Programming Approach for Fair Classification

William Yang, Lawrence Livermore National Lab, Livermore, CA, United States, Chaoyue Zhao, Shuai Huang, Ryan Lin

Traditional binary classification algorithms are prone to producing unfair results that favor certain demographic groups over others. This inequity is often exacerbated in unbalanced datasets where the number of entries from a majority group significantly outweighs the entries from a minority group. We use a mixed-integer-programming (MIP) framework to formulate our Rawlsian fairness methohdology to address these inequities. Our methodology prioritizes the performance of the worst-off demographic group, and our specific formulation can produce interpretable solutions by directly optimizing sparsity. Additionally, it provides flexibility for users to achieve interpretable solutions in multiple ways.

We directly optimize accuracy and sparsity while providing model flexibility that enhances interpretability by allowing the user to adjust sparsity through penalization in the objective or with upper-bounding constraints. We also offer flexibility to obtain real or integer coefficients based on domain-based preferences. Furthermore, our model outputs a single optimal threshold, which alleviates the user from selecting a threshold value before implementing the model. Consequently, our model provides deterministic classification results instead of probability values assigned to each individual.

Our experimental results display our Rawlsian MIP's ability to prioritize the performance of the worst-off group while still maintaining a high classification accuracy overall, even when performed on an unbalanced dataset. We also showcase the flexibility of our method in its ability to adjust the sparsity and interpretability of the results. Lastly, we compare our results to other fairness metrics, and discuss our method's differences with respect to other Rawlsian classification methods.

5 - Mixture-of-Experts Modeling with Shape Constraints and Its Application to Materials and Industrial Systems

Yating Fang, Rutgers University, Industrial & Systems Engineering, Piscataway, NJ, United States, Qianqian Zhao, Ryan Sills, Ahmed Aziz Ezzat

Shape constraints, such as monotonicity, arise in many engineering and scientific applications, as motivated by some underlying physics or existing theory. We are interested in the problem of imposing monotonicity constraints in the context of mixture-of-experts (MoE) modeling – where multiple machine learning models (referred to as experts) are trained on different data subsets, and then merged -using internally estimated input-dependent weights (referred to as gates)- to produce a combined mixture model. The main motivation of MoE is to accommodate data heterogeneity arising from distinct physics-based regimes, and/or reduce computational cost of training large machine learning models. Unfortunately, the resulting mixture is not guaranteed to be monotone, even if the constituting experts are, limiting MoE's applicability to several scientific and engineering applications where monotonicity is desirable, if not necessary. In response, we propose novel variations to the MoE framework in order to enforce monotonicity on the resulting mixture, without sacrificing its predictive power. Our proposed method is shown to perform well on simulated datasets, as well as on real-world datasets from materials/industrial systems experiencing multi-physics regimes.

TA72

Regency - 703

Optimization and Surrogate Methods for Black-Box Systems

Invited Session

Data Mining

Chair: Nazanin Nezami, University of Illinois Chicago, Chicago, IL, United States

Co-Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Building Trust in Black-box Optimization: A Comprehensive Framework for Explainability

Nazanin Nezami, University of Illinois Chicago, Chicago, IL, United States

Optimizing costly black-box functions within a constrained evaluation budget presents significant challenges in many real-world applications. Surrogate Optimization (SO) is a common resolution, yet its proprietary nature introduced by the complexity of surrogate models and the sampling core (e.g., acquisition functions) often leads to a lack of explainability and transparency. While existing literature has primarily concentrated on enhancing convergence to global optima, the practical interpretation of newly proposed strategies remains underexplored, especially in batch evaluation settings. We propose Inclusive Explainability Metrics for Surrogate Optimization (IEMSO), a comprehensive set of model-agnostic metrics designed to enhance the transparency, trustworthiness, and explainability of the SO approaches. Through these metrics, we provide both intermediate and post-hoc explanations to practitioners before and after performing expensive evaluations to gain trust. We consider four primary categories of metrics, each targeting a specific aspect of the SO process: Sampling Core Metrics, Batch Properties Metrics, Optimization Process Metrics, and Feature Importance. Our experimental evaluations demonstrate the significant potential of the proposed metrics across different benchmarks.

2 - Bayesian Optimization for Inverse Problems with Intractable Likelihoods

Peter Frazier, Cornell / Uber, Ithaca, NY, United States

We consider calculation of the posterior density over uncertain parameters when the forward model is time-consuming to evaluate. This problem arises, for example, in epidemic modeling, where we wish to estimate parameters in an epidemic simulation by comparing its predictions to observed data, and where each simulation requires significant computation.

Bayesian optimization of composite functions (Astudillo & Frazier 2019) provides an efficient way to maximize the posterior density: use a Gaussian process to predict the output of the forward model at unseen parameter values; use knowledge of the likelihood and the prior to predict the unnormalized posterior; and use a classical acquisition function to choose the next point at which to evaluate the forward model. While quite efficient, this approach is limited to providing point estimates of the uncertain parameters.

We show how a significant generalization of these ideas supports approximate fully Bayesian inference over the uncertain parameters. By using a compositional Gaussian process model and an acquisiton function designed to prioritize localization of areas where the likelihood is high, we can draw samples from a high-quality approximate posterior with a small number of forward model evaluations. We illustrate these ideas with examples from epidemic modeling and climate-responsive plant breeding.

3 - Simulation Optimization with Non-Stationary Streaming Input Data

Haowei Wang, National University of Singapore, Singapore, Singapore, Songhao Wang, Szu Hui Ng

Simulation optimization has become an emerging tool to design and analysis of real-world systems. In stochastic simulation, input distribution is a main driving force to account for system randomness. Most existing works on input modeling focus on stationary input distributions. In reality, however, input distributions could experience sudden disruptive changes due to external factors. In this work, we consider input modeling through non-stationary streaming input data, where the input data arrive sequentially across different decision stages. Both the parameters of the input distributions and the disruptive change points are unknown. We use a Markov Switching Model to estimate the non-stationary input distributions, and design a surrogate based approach to solve the following optimization problem. The proposed surrogate and optimization algorithm can utilize the simulation results from all the past stages. A numerical study on an inventory system shows that our algorithm can solve the problem more efficiently compared to common approaches.

4 - Clustering Simulation Output Distributions for pre-optimization and online monitoring

Mohammadmahdi Ghasemloo, Texas A&M University, College Station, TX, United States, David Eckman

Analyzing simulation outputs using statistical learning methods can improve decision-making by exploring relationships between simulation inputs and outputs in preparation for optimization. We focus on the task of clustering multivariate distributions of simulation outputs to identify patterns and tradeoffs and to better understand the behavior of a system. We introduce a novel agglomerative clustering algorithm that leverages the regularized Wasserstein distance to group system configurations having similar output distributions. We explore several use cases for this framework, including anomaly detection and pre-optimization. In numerical experiments of a simulated call center, we apply the proposed methodology to identify staffing plans that achieve similar performance and to develop policies for intervening when the queue lengths indicate that system performance is at risk of sharply deteriorating.

TA73

Regency - 704

Advances in Machine Learning and Decision Analytics

Invited Session

Data Mining

Chair: Shouyi Wang, The University of Texas at Arlington, Arlington, TX, 76019, United States

Co-Chair: Dervis Ozay, The University of Texas at Arlington, Arlington, TX, 76155, United States

1 - Does YouTube affect medical decision making? Causal machine learning for health disparities

Minh Nguyen, Michigan State University, East Lansing, MI, United States

Understanding causal relationships is crucial for making decisions, especially in some critical fields like health care. In these critical fields, understanding associative relationships might not be enough. This work focuses on estimating causal effects of accessing YouTube (or Dr.YouTube) on medical decision making. In order to do this, we employ a list of traditional econometric as well as current advanced causal machine learning methods for treatment effect estimation. We utilize the Health Information National Trends Survey (HINTS), which is a reliable data set of the American public’s knowledge of health-related information. The results show that Dr. YouTube consistently positively impacts medical decision making of the individuals. This result is robust under different refute tests. Moreover, we go beyond general causal results by investigating and providing new causal understanding on health disparities. This work implies that YouTube is not just a source for accessing medical information as the existing literature but also the one for medical decision making.

2 - Treatment Response Prediction of Lung Cancer Using Deep Learning

Shouyi Wang, The University of Texas at Arlington, Arlington, TX, United States

In this presentation, we explore the revolutionary capabilities of deep learning in predicting treatment responses for lung cancer patients. Utilizing a comprehensive dataset comprising clinical features, genomic data, and imaging studies, we developed a robust deep learning framework capable of assessing patient-specific treatment outcomes. Our model integrates convolutional neural networks (CNNs) for image analysis and recurrent neural networks (RNNs) for sequential data processing, achieving a significant predictive accuracy compared to traditional statistical methods. We also address the challenges of model interpretability and data heterogeneity, implementing techniques such as layer-wise relevance propagation to elucidate decision-making processes. The results demonstrate the potential of deep learning to personalize therapy plans, thus improving survival rates and quality of life for lung cancer patients. This talk will detail the methodologies, results, and future directions of our research in transforming oncological treatment paradigms.

3 - Improving Binary Classification Performance under Class Imbalance via Minority Sub-Classes

Tai-Jung Chen, Virginia Tech, Blacksburg, VA, United States, Kwokleung Tsui

Many real-world classification tasks are highly class imbalance. This results in the machine learning classifiers performing poorly on the minority class samples although performing well within the majority class samples. Given the sub-class information within the minority class, we found out local models, which are trained on majority class samples and each minority sub-class samples respectively, had higher precision and specificity due to lower false positive classifications. We propose DNC, (Divide aNd Conquer), a method that utilizing the minority sub-class information to divide the imbalance binary classification problem into several sub-problems, conquer each sub-problem via local models one by one, and aggregate the classifications from the local models to enhance the performance of the original imbalance binary classification. We showed that DNC can improve precision comparing to using the conventional binary classification method. While

DNC and One-versus-One (OvO) have similar performance, DNC is more time efficient and can reduce the number of classifiers trained from k chose 2 to $k - 1$ (where k is the number of classes)*, due to the design of training fewer local models.*

4 - Randomization Against Non-Stationarity: A Best Arm Identification Algorithm for Smart Healthcare

Yanwen Li, Nanyang Technological University, Singapore, Singapore, Zhenzhen Yan, Guangyan Gan

Policymakers for public healthcare services seek the strategy with the maximum prevention effectiveness of an infectious disease (e.g., HIV, COVID-19) among a set of prevention strategies. Clinical and simulation studies are the primary experimental methods to evaluate the effectiveness of prevention strategies, whereas the inherent non-stationarity (e.g., distributional shift) of the unpredictable bias between the experimental effectiveness and the true effectiveness of each strategy may cause non-negligible interference in the effectiveness evaluation and policy-making. In addition, limited experimental resources urge policymakers to improve the quality of experimental design and policy-making. This gives rise to a pressing need for an efficient resource allocation algorithm to properly evaluate each strategy and correctly identify the strategy with maximum effectiveness. We formulate the problem of disease prevention in homogeneous populations as a best arm identification problem in non-stationary stochastic environments, focusing on a fixed-budget scenario, where the strategies can be evaluated over a fixed number of rounds. Specifically, the quantitative results of the experimental effectiveness of each strategy are distorted versions of its true effectiveness, with the distortion process unbeknownst to policymakers. The goal is to design an allocation algorithm that maximizes the asymptotic decay rate of the probability of incorrect identification. To provide a systematic way of experimental design and policy-making, we apply the combination regimen of the large deviations theory and an adaptive randomization scheme to quantify a worthwhile tradeoff of resource allocation between effectively evaluating prevention strategies and actively defending against non-stationarity. Numerical results show the efficiency of our randomized algorithm.

TA74

Regency - 705

Energy Systems Integration

Invited Session

ENRE: Energy-Climate

Chair: Dharik Mallapragada, MIT Energy Initiative, Massachusetts Institute of Technology, 5 Metrotech Center, Cambridge, MA, 11201, United States

1 - Energy Transition Planning Using Stochastic Programming and a Control-Inspired Approach

Molly McDonald, Princeton University, Princeton, NJ, United States, Christos Maravelias

Energy system models can facilitate our understanding of how investment choices today can affect the ability to meet desired goals, such as reaching net zero emissions while still satisfying energy demand. However, although energy system models have been useful in demonstrating the potential impact of climate policy, most such models have two key limitations: 1) they are based on the assumption that the values of key parameters throughout the time horizon are deterministically known*;* and 2) decisions over the full planning horizon, which may be decades, are made at a single point in time at the beginning of the horizon. We propose different methods to address these limitations and thereby better represent the decision-making process in the presence of uncertainty. We use multistage stochastic programming (MSSP) to account for uncertainty in key parameters, such as future energy demands. We pair this method with a sequential decision-making approach that better represents the ability of a decision-maker to update decisions as uncertainties unfold. We then apply a simulation-based framework to evaluate the quality of the energy transition pathways that are developed by the proposed methods. We find that accounting for uncertainty with MSSP and having the ability to react to uncertainty with the sequential decision-making approach leads to more diverse energy transition pathways that are less sensitive to uncertainties in key parameters.

2 - Locational Marginal Pricing of Energy in Pipeline Transport of Natural Gas and Hydrogen with Carbon Offset Incentives

Mo Sodwatana, Stanford University, Stanford, CA, United States, Saif Kazi, Kaarthik Sundar, Adam Brandt, Anatoly Zlotnik

We propose an optimization formulation for locational pricing of energy transported through a pipeline network that carries mixtures of natural gas and hydrogen from distributed sources to consumers. The objective includes the economic value provided by the pipeline to consumers of energy and suppliers of natural gas and green hydrogen, as well as incentives to lower carbon emissions by consuming the latter instead of the former. The optimization is subject to the physics of gas flow and mixing in the pipeline network as well as engineering limits. In addition to formulating this mathematical program, we synthesize the Lagrangian and derive analytical expressions for the dual variables. We propose that the dual solution can be used to derive locational marginal prices of natural gas, hydrogen, and energy, as well as the decarbonization premium paid by consumers that receive hydrogen. We derive several properties of solutions obtained using the proposed market mechanism, and demonstrate them using case studies for standard 8-node and 40-node pipeline test networks. Finally, we show that optimization-based analysis of the type proposed here is critical for making sound decisions about economic policy and infrastructure expansion for blending green hydrogen into existing natural gas pipelines.

3 - Co-Optimization of Industrial, Fuel Supply, and Electricity Sector Models

Maxwell Brown, Colorado School of Mines, Golden, CO, CO, United States

I present on the methodology and implications of linking two high-resolution representations of the electricity and industrial sectors in a co-optimization framework. The results focus primarily on the differences of standalone versus coupled representations under decarbonization scenarios.

4 - Evaluating the impacts of technology and policy drivers on US energy system decarbonization

David Young, EPRI, Palo Alto, CA, United States, John Bistline, Geoffrey Blanford, Aranya Venkatesh

Policy drivers, such as the Inflation Reduction Act (IRA) and the EPA 111 power plant rules for existing and new sources, are expected to support decarbonization of the U.S. energy system and pathways to meet the U.S. Nationally Determined Contribution targets. How emerging technologies may evolve will also strongly impact decarbonization trajectories. In this study, we use EPRI's U.S. Regional Economy, Greenhouse Gas, and Energy Model (US-REGEN) framework to develop scenarios illustrating how the U.S. electric sector may respond to a range of policy and technology drivers. These include uncertainties in the technoeconomic performance and infrastructure constraints of emerging technologies such as carbon capture and storage and hydrogen, as well as sensitivities around IRA incentives and the 111 rules. US-REGEN includes a detailed dispatch and capacity expansion model of the electric sector, along with additional models that represent U.S. energy end-use and fuel production, conversion, and transportation. For each scenario, we examine and compare impacts on electric capacity, generation, and CO2 emissions, and highlight difference in regional outcomes.

5 - Competitiveness of underground thermal energy storage in electricity system

Alicia Wongel, Carnegie Science, Stanford, CA, United States, Austin Vernon, Ian McKay, Ken Caldeira

Thermal energy storage can help shift demand peaks in electricity systems. Different forms of energy storage may play an increasingly important role in reliably meeting electricity demand, helping to shift electricity from times of abundant generation to times of high residual demand. However, most storage technologies currently available come at a high cost. In this study, we assess under which conditions underground thermal storage with low energy cost could become cost-effective in electricity systems reliant on natural gas or on wind and solar generation. We find that in electricity systems reliant on natural gas, underground thermal storage could play a limited role in helping meet seasonal demand peaks, with some saving in natural gas capacity and fuel costs; in electricity systems reliant on wind and solar generation, underground thermal storage could play a substantial role in cost-effectively reducing wind and solar capacity required to satisfy loads thus reducing overall system cost.

TA75

Regency - 706

Advances in Sustainable Operations From Social and Environmental Perspectives

Invited Session

ENRE: Environment and Sustainability

Chair: Osman Alp, University of Calgary, Calgary, AB, Canada

1 - Self-Assessments for Supplier Sustainability: An Empirical Analysis

Tarkan Tan, University of Zurich, Zurich, Switzerland, Vrushali Patil, Shaunak Dabadghao, Sonja Rispens, Evangelia Demerouti

Large companies are increasingly accountable for the social and environmental impact of their suppliers. To ensure compliance, many companies implement sustainability standards and conduct audits. However, issues such as audit fraud, bribery, and fatigue undermine the accuracy of these audits. As a result, firms are shifting from a policing to a collaborative approach with suppliers, focusing on supplier self-assessment (SA). For instance, Philips BV requires suppliers to provide structured SAs with documentary evidence, promoting a trust-based rather than punitive approach. This method reportedly improves visibility of sustainability issues, engages suppliers more effectively, and enhances sustainability over time. Other companies like Volvo, Aldi, Home Depot, and Sanofi also adopt similar approaches. Despite these benefits, the accuracy of SAs remains a concern. Suppliers might manipulate their responses to appear more favorable to buyers. Factors contributing to inaccuracies include lack of time, resources, or understanding of the assessment parameters. Inaccurate SAs hinder buyers from mitigating sustainability risks and making informed improvement actions. In this talk we focus on enhancing the effectiveness of the SA approach considering these inherent inaccuracies. We analyze SA scores from 120 of Philips' Chinese suppliers (2016-2021) and compare them with validated (VAL) scores obtained through onsite visits. Using multi-linear regression analysis and t-tests, we investigate which sustainability dimensions have more accurate SAs, whether accuracy improves with multiple SA cycles, and how supplier characteristics like size, country, and relationship strength affect accuracy. Our goal is to help buyer firms predict inaccuracies and better target their validation and improvement efforts.

2 - Socially Efficient Rooptop Solar Energy Policies

Osman Alp, University of Calgary, Calgary, AB, Canada

In this talk, we consider alternative approaches for rooftop solar energy adoption that maximizes the benefit to various stakeholders in the society.

3 - Managing Product-Reusability Under Supply Disruptions

Prashant Chintapalli, Ivey Business School, Western University, London, ON, Canada, Kumar Rajaram, Nishant Kumar Verma

We model and analyze product reusability, executed through refurbishment, amidst potential supply disruptions. Consumers can trade in used units, which can later be refurbished and sold. Using a three-period model, we determine the optimal degree of product reusability, trade-in and refurbishment policies, trade-in fee, and the prices of new and refurbished units chosen by the firm.

4 - Who Leads the Way? Buyer vs. Supplier Initiatives in Supply Chain Carbon Footprint Reduction

Elif Kuscu, University of Zurich, Zurich, Switzerland, Ozgen Karaer, Tarkan Tan

We study two supply chain models: (i) Buyer-Led Supplier Development and (ii) Supplier-Led Supplier Development, which are consisting of a buyer and a supplier. We analyze which player should initiate the carbon footprint reduction efforts in the supply chain under a scope 3 emission tax by using different levers. The supplier, as the producer, may invest to reduce the product carbon footprint. The buyer, as the downstream partner, may incentivize the supplier through supplier development initiatives. In the buyer-led model, the buyer offers a wholesale price premium rate contingent on improvement, and the supplier decides her carbon footprint reduction efforts accordingly. In the supplier-led model, the supplier offers sharing the carbon footprint reduction investment cost and the buyer decides the improvement level. We compare these two models and gain insights into their effectiveness under varying market conditions. Our results show that carbon footprint reduction is guaranteed if the supplier takes the lead, however, a wider range of improvement opportunities occur when the buyer initiates the reduction effort.

TA76

Regency - 707

Industrial Decarbonization III: Computational Perspectives for Promoting Industrial Decarbonization

Invited Session

ENRE: Other Energy

Chair: Paritosh Ramanan, Oklahoma State University, Stillwater, OK, United States

Co-Chair: Paritosh Ramanan, Oklahoma State University, Stillwater, OK, United States

Co-Chair: Zheyu Jiang, Oklahoma State University, Stillwater, OK, 74078, United States

1 - Decarbonization of Steam Cracking Units with joint Power System Planning

Steam cracking is the most prevalent petrochemical process that uses high temperature heat to break down hydrocarbons, such as natural gas and naphtha, into more valuable unsaturated hydrocarbons such as ethylene and propylene, which are the building blocks of polymers and other specialty chemicals. Steam cracking units around the world produce over 150 million tons of ethylene and propylene every year. Due to its endothermic nature, significant heat duty is required for steam cracking and downstream processes. It is estimated that steam cracking itself accounts for 8% of the total primary energy in chemical industry. As chemical companies are setting ambitious sustainability goals and greenhouse gas emission reduction targets, active research and development is undergoing to replace highly fossil intensive cracking furnace with electrified furnace, which can be powered by (variable) renewable electricity (VRE). For example, in June 2022, Dow Chemical and Shell announced its startup of the first experimental electrified cracker unit in Amsterdam. These ongoing trends in decarbonizing steam cracking via electrification establish unique opportunities and challenges from an operational perspective. First, it is expected that conventional crackers will still the abundant in the foreseeable future, as it is capital intensive to simply replace existing heat-driven crackers with new electrified crackers. Second, due to the installation of electrified crackers, chemical plants will become microgrids that possess a diverse energy portfolio, ranging from conventional fossil-based heat to electricity as well as other energy carriers such as hydrogen. Thus, operational challenges emerge as electrified chemical plants need to be integrated with the power systems. To systematically address these complications, in this talk, we present a comprehensive modeling and optimization framework that coordinates the operation of both conventional heat-driven crackers and electrified crackers with power systems that are becoming more intermittent due to increasing VRE proportions. First, we develop a first principle model to quantify the power demand of conventional and electrified cracking, subject to specific feedstock constraints as well as purity and yield targets. The resulting differential algebraic equation (DAE) based model is then incorporated into an optimization framework to determine the minimum heat duty/power requirement, which is subsequently embedded in an optimization formulation that encompasses the unit commitment of connected power systems. To tackle the computational challenges associated with the resulting a large-scale MILP problem, we adopt Bender's decomposition and partition the original problem into a master problem (unit commitment) and subproblems (microgrids). The microgrid subproblems, which are themselves MILPs due to the hourly decision making required for multiple energy conversion and storage units involved, undergo relaxation of binary variables followed by primal recovery to introduce integrality cuts into the master problem. This decomposition approach enables efficient optimization by exploiting the structure of the problem and reducing computational complexity. Our case study of steam cracking electrification in Texas demonstrates the tractability of the proposed modeling and optimization framework and shows that significant greenhouse gas emission reduction can be achieved.

2 - Importance Sampling in Variational Autoencoders to generate Industrial Carbon Emissions Scenarios

Huynh Quang Nguyen Vo, Oklahoma State University, Stillwater, OK, United States, Paritosh Ramanan

Concerns about the negative environmental impact of fossil fuels have spurred significant interest in decarbonization efforts. Renewable energy remains a key focus in these efforts but faces numerous challenges due to uncertainties in supply and demand. Stochastic programming provides a powerful framework for decision-making under such uncertainties. A key technique to effectively solve these problems is scenario generation, which organizes stochastic process outcomes into finite discrete distributions, facilitating mathematical programming models. The solutions to stochastic programs are heavily influenced by the quality of the generated scenarios, whose fidelity being a significant concern in decentralized applications. A major issue is the likelihood of scenario occurrence. In this talk, we explore the potential of variational auto-encoders (VAEs) and their decentralized variants in importance sampling, a method for computing mathematical expectations with respect to a target distribution - specifically, the scenario distributions - to address this challenge effectively.

3 - Decarbonization of Chemical Process Industries via Electrification

Zheyu Jiang, Oklahoma State University, Stillwater, OK, United States, H M Mohaimanul Islam

In 2018 the US manufacturing sector used ~20% of the country's primary energy and emitted ~18% of its total greenhouse gases (GHGs). The top two US energy-use processing industries, namely the chemicals and petroleum refining industries, use nearly half of the manufacturing sector's primary energy and emit half of its GHGs. The majority of the process energy use in these sectors is for process heating. Currently, almost all this process heating is either met by the direct combustion of fossil fuels or through the use of steam that is generated directly or indirectly through the combustion of fossil fuels. Historically, for process heating applications, there has been a choice whether to use heat from the combustion of fossil fuels or to convert the combustion heat to electricity and then use the generated electricity. In this case, the extra cost and efficiency loss due to conversion to electricity does not really stand out compared to using combustion heat directly. However, as the US energy landscape continuous to shift toward decarbonized electricity, mostly solar and wind, the energy can be directly harvested as electricity, which makes it worth examining the potential impact of replacing combustion heat with electricity. Through simple calculations, we show that if all this heat was to be provided by resistive electrical heating, this would translate to roughly 36% of the U.S. electricity generation. Such a large power demand for just heating applications will put a huge strain on electric power infrastructure.

4 - Decentralized Operations Planning of Decarbonized Chemical Plants with Renewable-driven Transmission Systems

Richard Reed, Oklahoma State University, Stillwater, OK, United States, Paritosh Ramanan, Saba Ghasemi Naraghi, Zheyu Jiang

As the electrification of industrial process heating becomes common it is increasingly important to integrate clean energy from power systems into chemical plant processes. However, integrating private chemical plant operations with the public power grid raises privacy concerns. Decentralizing large-scale unit commitment in power systems and electrified chemical process heating in a mixed-integer problem can maintain this privacy, encouraging stakeholders to adopt renewable energy use, thereby aiding decarbonization efforts. In this talk we will discuss decentralized optimization techniques that implement iterative ADMM to reach a consensus. We demonstrate that by limiting shared information through decentralization, it is possible to satisfy demand needs of the chemical plant while maintaining the privacy of its' operations. This approach eliminates the need to move data while considering operational constraints of the power system and chemical plant, while procuring expected results and retaining the privacy of operations.

TA77

Regency - 708

Decomposition Methods in Network Optimization

Invited Session

Computing Society

Chair: Robert Curry, University of Arkansas, Fayetteville, AR, United States

1 - Mixed-Integer Programming Models for Solving Semi-Cooperative Attack Graphs Problems

Phanuel Allaissem Beremadji, University of Arkansas, Fayetteville, AR, United States, Robert Curry

In this paper, we investigate the Semi-Cooperative Attack Graph (SCAG) problem, focusing on optimizing attack strategies in interdependent infrastructures. We propose two mixed-integer programming (MIP) models to minimize the time required to complete attacks, accounting for shared resources and budget constraints among multiple attackers.

The first MIP model (MIP-C) uses continuous time variables and binary variables to track the time each state is reached and by which attacker. The second MIP model (MIP-D) utilizes discrete time periods and binary variables to determine if an attacker reaches a state during a particular time period. Both models are enhanced with valid inequalities and strategic parameter settings to improve computational efficiency and strengthen the linear programming relaxation. We finally present a series of computational results to explore the efficacy of both modeling approaches.

The SCAG model can model attacks on systems across various domains, including wireless sensor network security and defense strategies for interdependent infrastructures. The models contribute to a more comprehensive understanding of attack dynamics and optimization in real-world scenarios by considering cooperative attack scenarios where multiple attackers leverage shared resources.

2 - Survival Signature Estimation Using Optimization and Monte-Carlo Simulation for K ≥ 3 Classes of Nodes on Two-Terminal Networks.

Md Sazid Rahman, University of Arkansas, Fayetteville, AR, United States, Kelly Sullivan

This research develops an efficient approach to estimating survival signatures for two-terminal networks with more than two classes of components. Recently, the survival signature has gained substantial attention in the literature on network reliability estimation due to its unique separability property, which enables passing the network topology information independent of the failure distribution of the components. Following recent results from the literature, estimating the survival signature by Monte Carlo simulation entails solving a multi-objective maximum capacity path problem on a two-terminal network in each replication. We adapt a multi-objective Dijkstra's algorithm from the literature to construct the set of non-dominated paths solving the multi-objective maximum capacity path problem for each replication of the Monte-Carlo simulation. We have carried out experiments on random two-terminal networks and grid networks with three, four, and five classes of components. In these experiments, our version of the multi-objective Dijkstra algorithm was compared against four benchmark algorithms and an improvement technique that prunes some paths to be explored in the multi-objective Dijkstra algorithm setting lower bounds on capacities. We compared the run-time of our approach with all these benchmark approaches and found that the multi-objective Dijkstra algorithm performs significantly better in most instances.

3 - Maximum Flow Stochastic Network Interdiction with Uncertain Restructuring

Farhatul Janan, Clemson University, Clemson, SC, United States, Thomas Sharkey, Emily Tucker

We study the max flow stochastic network interdiction problems with uncertain restructuring (MF-SNIP-UR). In order to capture the underlying uncertainty associated with MF-SNIP-UR, we focus on the instances where interdictions may not be successful and where restructuring after successful interdictions happen stochastically. We propose a two-stage stochastic model (2S-SM) for this problem where the attacker makes the interdiction plan in the first stage, the network stochastically restructures based on the realization of successful interdictions, and then solves a maximum flow problem in the resulting network for the scenario. In addition, we capture the practical considerations that planned interdictions and then the resulting successful interdictions for each scenario should follow climbing the ladder constraints. We formulate the 2S-SM problem as a bi-level mixed integer problem (MIP) and then we convert it to a single-level MIP using the properties about restructuring and standard duality concepts. We examine sample average approximation (SAA) approaches to solve the single-level formulation of 2S-SM problem. We propose a fast approximation of the full 2S-SM by examining partial network formulations, which help find a valid lower and upper bound on the full network. This approximation can solve problems that have at least four times the number of scenarios as the original SAA approach. We demonstrate the importance of our problem by examining applications in drug trafficking networks.

4 - Parcel Truckload Trucking: Solving UPS's Large-Scale Network Design Challenge

Belgacem Bouzaiene-Ayari, UPS, Parsippany, NJ, United States

UPS possesses the world's largest integrated parcel truckload (PT) network, spanning across multiple service types and supported by an extended set of center and hub locations and a sophisticated multimodal transportation system featuring an arrays of equipment types. Both strategic hub placement and demand routing are pivotal factors in establishing such a network that fulfills service commitment under various operational constraints, while minimizing the total operational cost. This cost is composed of two components: the transportation expenses and the package handling cost incurred at the consolidation hubs. The network configuration must consider several other constraints revolving around: (a) equipment and driver management, (b) capacity of the sorting operations conducted at the various network hubs, (c) restrictions on feasible transportation modes by service type, and (d) the number of inbound and outbound doors available at the network facilities.

In this study, we formulate the problem of creating an efficient network configuration as a network design problem. Our model is defined on a compact space to ensure scalability. Additionally, we propose a novel solution approach based on well-established decomposition techniques such as column generation, benders decomposition, and Lagrange relaxation. Under the assumption of consistent demand throughout the planning horizon, the model and its algorithmic approach have been proven to deliver high-quality solutions for large-scale problem scenarios encompassing up to five thousand sorting operations, four million arcs, and two million origin-destination demand pairs.

TA78

Regency - 709

Redistricting

Invited Session

Computing Society

Chair: Soraya Ezazipour, Oklahoma state university, Stillwater, OK, United States

Co-Chair: Maral Shahmizad, Oklahoma State University, Stillwater, OK, United States

1 - Political districting to maximize whole counties

Maral Shahmizad, Oklahoma State University, Stillwater, OK, United States

When drawing contiguous, population-balanced political districts, how many counties can be kept whole, i.e., not split across multiple districts? To answer this question, we propose a combinatorial Benders decomposition algorithm. The main problem decides which counties to keep whole, and the subproblem coarsens the selected counties and attempts to find a feasible plan. Key to the approach's success is a new class of valid inequalities. We apply the approach to all congressional and legislative districting instances across the USA and find provably optimal solutions for most of them. Our code and results are available on GitHub.

2 - Political districting to optimize the Polsby-Popper compactness score

Soraya Ezazipour, Oklahoma state university, Stillwater, OK, United States

In the academic literature and in expert testimony, the Polsby-Popper score is the most popular way to measure the compactness of a political district. Given a district with area A and perimeter P, its Polsby-Popper score is given by $(4 \pi A)/P^2$. This score takes values between zero and one, with circular districts achieving a perfect score of one. In this paper, we propose the first mathematical optimization models to draw districts with optimum Polsby-Popper score. Specifically, we propose new mixed-integer second-order cone programs (MISOCPs), which can be solved with existing optimization software. We investigate their tractability by applying them to real-life districting instances in the USA. Experiments show that they are able to: (i) identify the most compact majority-minority districts at the tract level; (ii) identify the most compact districting plans at the county level; and (iii) refine existing tract-level plans to make them substantially more compact. Our techniques could be used by plaintiffs when seeking to overturn maps that dilute the voting strength of minority groups. Our python code and results are publicly available on GitHub.

3 - Local Walks and Network Partitioning for Discrete Redistricting Problems

Daryl DeFord, Washington State University, Pullman, WA, United States

Many redistricting problems, including drawing electoral boundaries in the USA, are naturally modeled as graph partitioning problems, which allows for analysis using discrete optimization and sampling methods. Applying this approach successfully in practice requires both operationalizing legislative text and exploring complex Pareto frontiers and in recent years spanning tree-based approaches have become popular for generating large samples of comparison plans. In this talk I will discuss recent applications and extensions of these techniques, including for court cases, support of line-drawing commissions, and evaluating nonpartisan justifications for proposed plans. Along the way I will present related open problems and proposals based on the cycle basis walk and classic network partitioning methods including spectral clustering and modularity.

Tuesday, October 22, 9:45 AM - 10:35 AM

Summit - Ballroom 1

Preparing Interdisciplinary Leaders: Re-Imagine Your Department As A Place For Educating Future Operations Research Leaders Who Can Tackle Complex And Nuanced Problems

Plenary/Keynote Session

Plenary

Chair: Julie Ivy, University of Michigan, Ann Arbor, MI, United States

1 - Preparing Interdisciplinary Leaders: Re-Imagine Your Department As A Place For Educating Future Operations Research Leaders Who Can Tackle Complex And Nuanced Problems

Harriet Nembhard, Harvey Mudd College, Claremont, CA, United States

In the face of today's multifaceted global challenges, the role of Operations Research (OR) is more critical than ever. This plenary will explore how academic departments can be transformed into incubators for future OR leaders who are adept at addressing complex societal issues through interdisciplinary approaches. Beginning with a personal narrative, the talk highlights the urgent need for integrating diverse perspectives in OR education and practice. Drawing from successful initiatives at Harvey Mudd College and real-world applications, we discuss strategies for embedding interdisciplinary thinking, enhancing curricula, and fostering collaboration with industry and community partners. Participants will gain actionable insights on aligning academic programs with strategic goals, promoting effective leadership, and preparing students to create impactful, innovative OR solutions that contribute to achieving the UN Sustainable Development Goals (SDGs). This session aims to inspire and equip OR educators and practitioners to reimagine their departments as dynamic environments for training the next generation of leaders capable of making smarter decisions for a better world.

Tuesday, October 22, 10:45 AM - 12:00 PM

TB01

Summit - 320

Data Analytics and Optimization for Smart Industry

Invited Session

Scheduling and Project Management

Chair: Lixin Tang, Northeastern University, Shenyang, N/A

1 - A Branch-and-Price Algorithm for Two-Stage Cutting Stock Problem based on Triple Transfer and One Feedback

Ying Meng, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, 110819, Shenyang, China, People's Republic of, Lixin Tang, Roberto Baldacci, Md AI Amin, Jiyin Liu, Marco A. Boschetti

The classical cutting stock problem (CSP) involves efficiently producing copies of items with different weights and demands using a minimal number of bins while ensuring that no bin exceeds its weight capacity. Various versions of CSPs have emerged to address specific constraints and objectives in different applications. This presentation focuses on a new one-dimensional two-stage CSP with flexible item sizes, known as the one-dimensional two-stage cutting stock with flexible item sizes problem (TSCSP), which arises in the steel industry. In the problem, each order is associated with a width range and a total demand. For the TSCSP, we develop the pattern-based mathematical formulation, and propose some valid inequalities to improve its linear programming relaxation. Then, an exact algorithm within a branch-and-price (BP) framework is developed to solve the problem optimally. In the exact algorithm, a novel label-setting-based pricing algorithm is designed for solving the pricing subproblems. A primal heuristic that relies on a depth-first heuristic search in the BP tree is also designed to speed up the solution process. We validate the method's performance through extensive computational experiments on benchmark instances from literature and real-world scenarios, confirming its effectiveness in solving TSCSP instances with up to 150 orders.

2 - Combinatorial Benders Approach for Quay Crane Scheduling Problem based on Triple Transfer and One Feedback

Defeng Sun, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang, Roberto Baldacci

In this paper, we propose an exact decomposition approach to solve the Quay Crane Scheduling Problem in container terminals. The problem is decomposed into two easier problems, i.e. a routing master problem and a scheduling sub-problem. The master problem probes possible routing solutions for quay cranes while the sub-problem determines the completion times of each task and each quay crane. Several types of valid inequalities are added into the master problem to significantly reduce the search space and improve the lower bound. Precedence-based combinatorial cuts and multiple cuts are generated to further accelerate the algorithm convergence with less computational effort. Computational tests on the benchmark instances verify the effectiveness of the proposed approach.

3 - Generalized Disjunctive Programming Modeling and Reinforcement Learning-based Optimization Method for Oxygen Production Scheduling based on Triple Transfer and One Feedback

Shengnan Zhao, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang

As one of the important energy media in steel production, oxygen production has the characteristics of energy-intensive and high emission. Steel companies can take advantage of optimal schedule of oxygen production to reduce energy cost and emission. In this work, we addressed an optimal scheduling problem of oxygen production in steel company with consideration of variable load of the oxygen production units. The problem is formulated as a generalized disjunctive programming (GDP) model and reformulated with hull reformulation method. The problem is then solved by a novel Lagrangian relaxation algorithm, in which a Q-learning method is designed to scientifically set the initial step size coefficient. Numerical experiments show that the proposed strategy can effectively improve the convergence speed.

4 - Data Circulation Mechanism and Technology to link Steel and Equipment Manufacturing Industries base on Triple Transfer and One Feedback

Gongshu Wang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, 110819, China, People's Republic of, Lixin Tang

The steel industry plays a vital role in supplying raw materials to the equipment manufacturing sector, while the equipment manufacturing industry reciprocates by producing essential machinery serving the steel sector. Together, they form a Manufacture-Circulation Industrial System (MCIS) characterized by circularity. The industrial Internet serves as a critical infrastructure for connecting these industries in the digital realm. However, challenges arise due to data ownership issues, hindering data sharing and circulation. To address this, we've integrated various technologies like identity parsing, blockchain, and privacy computing to establish a robust data circulation mechanism.

Through identity parsing, unique identifiers are assigned to physical entities and virtual objects, ensuring traceability. Blockchain technology underpins a secure and orderly data circulation system, enhancing trustworthiness. Furthermore, privacy-enhancing technologies safeguard data security. Our goal is to facilitate dynamic data flows across enterprises by leveraging key technologies for data circulation.

5 - Modeling and Solution to Production and Inventory Planning Problem based on Triple Transfer and One Feedback

Yang Yang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang

The production process of steel enterprise includes multiple production stages, and there is work-in-process inventory between two consecutive stages. Thus, the decision of production and inventory volume are closely related. Moreover, the production process inevitably involves logistics operation and energy consumption, and the information technology enables the information sharing among the different production stages. This talk studies a production and inventory planning problem based on triple transfer (production, logistics, energy transfer) and one feedback (information feedback). To describe the problem, a mixed integer programming model is formulated. Then a CG-BS algorithm, combining Column Generation (CG) and Beam Search (BS), is developed to solve the problem. Finally, the effectiveness of the algorithm is verified through the computational experiments.

TB02

Summit - 321

Computational/Statistical Methods for Uncertainty Quantification: Part I

Invited Session

Quality, Statistics and Reliability

Chair: Ozge Surer, Miami University, Oxford, OH, United States

Co-Chair: Moses Chan, Northwestern University, Evanston, IL, United States

1 - Active Learning for Multi-Fidelity Computer Experiments with Adaptive Non-Stationary Kernel Function

Romain Boutelet, Michigan State University, East Lansing, MI, United States, Chih-Li Sung

Simulating complex physical processes across a domain of input parameters can be very computationally expensive. Multi-Fidelity Surrogate Modeling can resolve this issue by integrating cheaper simulations with the expensive ones in order to obtain better predictions at a reasonable cost. We are specifically interested in computer experiments that involve the use of Finite Element Methods with a real-valued tuning parameter that determines the fidelity of the numerical output. In these cases, integrating this fidelity parameter in the analysis enables us to make inference on fidelity levels that have not been observed yet. Such models have been developed and we propose a new Adaptive Non-stationary Kernel Function which more accurately reflects the behavior of computer simulation outputs. In addition, we aim to create a sequential design based on the Integrated Mean Square Prediction Error (IMSPE) to identify the best design points across input parameters and fidelity parameter, while taking into account the computational cost associated with the fidelity parameter. We illustrate this methodology through synthetic examples and applications to finite element analysis.

2 - Gaussian Processes with Sparse Grids

Haoyuan Chen, Texas A&M University, College Station, TX, United States, Rui Tuo

Gaussian processes (GPs) are widely used in non-parametric Bayesian modeling, and play an important role in various statistical and machine learning applications. In a variety tasks of uncertainty quantification, generating random sample paths of GPs is of interest. As GP sampling requires generating high-dimensional Gaussian random vectors, it is computationally challenging if a direct method, such as the Cholesky decomposition, is used. In this paper, we propose a scalable algorithm for sampling random realizations of the prior and posterior of GP models. The proposed algorithm leverages inducing points approximation with sparse grids, as well as additive Schwarz preconditioners, which reduce computational complexity, and ensure fast convergence. We demonstrate the efficacy and accuracy of the proposed method through a series of experiments and comparisons with other recent works.

3 - Diverse Expected Improvement (DEI): Diverse Bayesian Optimization for Expensive Computer Simulators

Simon Mak, Duke University, Durham, NC, United States

The optimization of expensive black-box simulators arises in a myriad of modern scientific and engineering applications. Bayesian optimization provides an appealing solution, by leveraging a fitted surrogate model to sequentially guide simulator evaluations. In practical problems, however, the goal is often not to obtain a single good solution, but a "basket" of good solutions from which users can choose for downstream decision-making. This arises in our motivating application for real-time control of internal combustion engines (ICEs), where a diverse set of control strategies are desired for robust and timely control. Despite its importance, there has been little work on this front for Bayesian optimization. We thus propose a new Diverse Expected Improvement (DEI) method, which extends the well-known Expected Improvement method to encourage diversity between "ϵ-optimal" solutions, i.e., solutions with objectives within a small ϵ from a global optimum. The DEI jointly targets two goals: the exploration of ϵ-optimal regions on the solution space, and the exploitation of promising solutions that may improve upon the current best solution. One advantage of the DEI is that it admits a closed-form acquisition function under a Gaussian process surrogate model, which facilitates efficient sequential queries via automatic differentiation. We demonstrate the improvement of DEI over state-of-the-art methods for diverse optimization in a suite of numerical experiments. We then explore the DEI in two practical applications, the first on rover trajectory optimization and the second for real-time control of ICEs.

4 - Simulation Experiment Design for Calibration via Active Learning

Ozge Surer, Miami University, Oxford, OH, United States

Simulation models often have parameters as input and return outputs to understand the behavior of complex systems. Calibration is the process of estimating the values of the parameters in a simulation model in light of observed data from the system that is being simulated. When simulation models are expensive, emulators are built with simulation data as a computationally efficient approximation of an expensive model. Instead of repeatedly running an expensive simulation model during the calibration process, an emulator then can be used to predict model outputs. Sequential design with an intelligent selection criterion can guide the process of collecting simulation data to build an emulator, making the calibration process more efficient and effective. This presentation focuses on two novel criteria for sequentially acquiring new simulation data in an active learning setting by considering uncertainties on the posterior density of parameters. Analysis of several simulation experiments and real-data simulation experiments from epidemiology demonstrates that proposed approaches result in improved posterior and field predictions.

TB03

Summit - 322

Tensor Data Analysis for Advanced Clustering and Process Monitoring

Invited Session

Quality, Statistics and Reliability

Chair: Jiuyun Hu, Arizona State University, Chandler, AZ, United States

Co-Chair: Hao Yan, Arizona State University, Tempe, AZ, United States

1 - Average vs Tail: Extreme Pattern Discovery of Metro Passenger Flows

Kai Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of

Modern metro systems generate sheer amounts of passenger travel records every day. While most works in transportation research are devoted to exploring the day-to-day variation of passenger mobility from these big traffic data, the variation is measured with respect to the mean or average of traffic flows. The interest of this work, by contrast, lies in the tail of traffic flow distribution, which accounts for the extreme level of traffic demands and the day-to-day variation of which is more relevant to risk control or capacity planning in metro operations. By characterizing the traffic flow tail behaviors by generalized quantile, where quantile and expectile are two special cases, and representing all traffic flows within a day as a higher order tensor, we propose a novel generalized quantile tensor decomposition to unveil extreme patterns that are associated with the day-to-day tail variation. Specifically, we consider an asymmetric loss function that assigns different weights to different elements of tensor, and then adopt it to decompose the daily traffic flow tensor into a series of rank-one basic tensors, each of which contains extreme patterns for multiple traffic modes (e.g., flow type, passenger type, time in a day) and is further induced to be sparse and smooth by regularization for better interpretation. An efficient learning algorithm is designed based on the least asymmetrically weighted squares and alternating direction method of multipliers. Our proposed methodology is finally applied to the Hong Kong metro dataset.

2 - Personalized Tucker Decomposition with Exponential Family

Hao Yan, Arizona State University, Tempe, AZ, United States

We propose a generalized personalized Tucker Decomposition to distinguish and decouple the shared and unique information in tensor format from different sources. The shared information is modeled by a global shared tensor and localized unique tensors. Several case studies including image monitoring and disease hotspot detection will be studied.

3 - Total Monitoring of Causation-Based Multivariate Processes

Xiyao Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Kaizong Bai, Jian Li

We consider monitoring multiple quality characteristics with causal relationships among them, and shifts occurring in cause variables may propagate to effect ones. Usually a Bayesian network (BN) can describe causation, where through directed arcs a variable depends on its direct cause variables. In the literature, some monitoring methods have been proposed based on BNs, but these approaches focus on merely the means of each variable. By fully exploiting the BN structure, we propose for each quality characteristic to monitor shifts in its mean and variance, as well as in the functional dependency on its direct cause variables, and therefore achieve total monitoring of causation-based multivariate processes. Specifically, we propose two causation-based control charts. The first chart serves as a general tool for process monitoring, while the second one incorporates directional information, enabling the identification of shift locations through a diagnostic prescription. Simulations have validated the effectiveness of the proposed monitoring and diagnostic schemes.

4 - Stream of Variation Modeling and Monitoring for Heterogeneous Profiles in Multi-Stage Manufacturing Processes

peiyao liu, Tsinghua University, Beijing, China, People's Republic of, Chen Zhang

Advanced manufacturing systems typically involve multiple operating stages, where in-process data are observed in the form of profiles that contain rich information for process monitoring. Emerging methodologies have been developed for multi-stage profile monitoring. However, the cases where different stages' profiles are heterogeneous and variations propagate over stages have not been addressed so far. To this end, this article proposes a heterogeneous profiles-based stream of variation (HP-SoV) framework for multi-stage manufacturing process monitoring. It uses functional decomposition to extract features of heterogeneous profiles from different stages, where the decomposition coefficients are regarded as latent states that propagate along consecutive stages to capture the variation propagation. HP-SoV includes many existing models as its special cases, and enjoys an efficient inference algorithm via maximum likelihood estimation. Based on the one-step-ahead forecast errors of HP-SoV, a group monitoring scheme is further developed for online monitoring. Extensive numerical studies explore the modeling robustness and monitoring effectiveness of HP-SoV under different settings. A case study demonstrates the applicability of HP-SoV in multi-stage process monitoring for heterogeneous profiles.

5 - GAN-based High-throughput Single Track Characterization to Reduce Variability and Experimental Costs in Laser Powder Bed Fusion Additive Manufacturing

Jiahui Ye, Texas A&M University, College Station, TX, United States, John Coleman, Gerald Knapp, Amra Peles, Chase Joslin, Alex Plotkowski, Alaa Elwany

Most process optimization methods in laser powder bed fusion additive manufacturing rely on single-track experiments as an efficient screening step. However, their variability due to melting pool instabilities is often overlooked. We propose a Generative Adversarial Network (GAN) based high-throughput single-track analysis method for rapid investigation of heteroscedastic variance in measurements while minimizing labor-intensive efforts. Coupling this method with physics-based defect criteria, we develop optimal process maps with uncertainty bands and further show how this approach easily applies to different processing conditions with minimal computational and experimental burden. This research was sponsored by National Science Foundation, the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Advanced Materials and Manufacturing Technology Office.

TB04

Summit - 323

Optimization and Interpretability in AI

Contributed Session

Chair: George Athanasopoulos, Monash University, Caulfield East, Australia

1 - Time-Varying Graph Learning for Unevenly Spaced Time Series Data

Masaki Takahashi, Graduate School of Information Science and Technology, Osaka University, Suita, Japan, Satoru Iwasaki

We propose a method for estimating the structure of time-varying graphs underlying time series data observed at unevenly spaced intervals at each vertex on the graph. We formulate the problem as a convex optimization problem, extending existing time-varying graph learning methods to unevenly spaced time series data without increasing the number of required hyperparameters. Additionally, we prove that both the existing method and our proposed method contain one unnecessary (not need to be tuned) hyperparameter. Moreover, we prove that the globally optimal solution of the formulated convex optimization problem can be obtained using the primal-dual splitting algorithm, and we provide an efficient algorithm. The effectiveness of our proposed method is also confirmed through several numerical experiments.

2 - Evaluation and Selection of Solution Procedures to Data-Driven Stochastic Programming: a Statistical Approach via Auxiliary Problems

Xianghua Gan, Southwestern University of Finance and Economics, Chengdu, China, People's Republic of

We investigate a data-driven stochastic programming (DDSP) challenge where the distribution of a random parameter from a predefined set is unknown and only inferred from observed samples. Existing literature presents various solution methods for DDSP, such as empirical optimization (EO), sample average approximation (SAA), operational statistics (OS), and distributionally robust optimization (DRO). Choosing the appropriate solution method is complex, especially with limited data. Each method offers a unique solution estimator that maps samples to real-valued solutions, with the operational risk of the DDSP problem depending on both the solution estimator and the random parameter. To estimate the solution estimator from an observed sample, we introduce auxiliary problems corresponding to each oracle distribution in the predefined set. Each auxiliary problem, mirroring the original, uses a sample from its specific oracle distribution. With infinitely many samples from an oracle distribution, we can precisely determine the true solution estimator and its operational risk for each auxiliary problem. The operational risk for the original problem is estimated using a weighted average of risks from these auxiliary problems, with weights based on the likelihoods of the oracle distributions derived from the observed sample. Additionally, we develop hypothesis tests to evaluate the performance of different solution procedures. We apply our methodology to a data-driven newsvendor problem, highlighting the acquisition of solution estimators, assessment of operational risks, and comparative analysis through hypothesis tests.

3 - Modelling Unstructured Information of External Events as Demand Sensing Signals using Generative AI

Belinda Soh, Singapore Institute of Technology, Singapore, Singapore, Shanshan Yang, Humza Akhtar, David Weidong Lin, Louiz Cheng Yin Lee, Wenda Chen, Ning Liu

Numerous studies have revealed a plateau in demand forecast accuracy in industry application. Demand planners tune the forecast accuracy based on their domain knowledge and understanding of market conditions. The accuracy of a forecast is thus subjected to human bias and the boundaries of the planner' experience. As the market grows more dynamic, comprehensively tracking market conditions manually becomes complex. Research and commercial applications have shown that inclusion of structured indicators such as weather to forecast demand models can be achievable. However, for market signals, they are largely unstructured data varying from quantitative to qualitative, and sometimes discrete. It is challenging for forecast models to comprehend such data. This paper makes first attempts to leverage on Large-Language Models to automate the knowledge acquisition and subsequent refinement into structured time-series data for input in forecasting models. In particular, the research will focus on real world events (i.e. disasters, civil/political conflicts, economic macrotrend) as input. The main contributions of the study are: (1) design for usage of vector database with Generative AI to process mined information on real world events into machine readable format; (2) development of a metrics framework to tackle the lack of standardized quantification of severity of real world events in existing demand sensing research (3) prove of significance of real world events' impact on demand based on correlation study using Time Lagged Cross Correlation and Dynamic Time Warping on the quantified indicator and real sales data taken from the M5 Competition.

4 - Optimizing Demand Forecasting for a Greener and Resilient Industry in the Competitive Market

Han Dong, Marriott International, Bethsda, MD, United States, Lokesh Merupula, Zhenzhen Zhu, Zelin Jiang

Due to the flexibility of return policies, high uncertainty, and competitions in the local market, traditional industries face significant challenges due to unpredictable demand, leading to inefficiencies and increased operational costs. This study presents an innovative bi-stage model for forecasting demand, integrating machine learning models with a two-stage approach to improve the accuracy, optimize resource allocation, and enhance resilience in the sector.

The model adopts a two-stage model structure. The first stage uses time series model to predict the demand in long term, while the second model refines these predictions by analyzing patterns in the demand over time to accommodate changes in short term.

The validation results of the model demonstrate its effectiveness in improving accuracy with uncertainties and reducing computing costs. Implementing this forecasting tool enables agents to make data-driven decisions, optimize production, and enhance inventory management, leading to more sustainable and efficient operations.

This research promotes sustainable practices in the industry. By reducing the uncertainty associated with demand, agents can better align operations with environmental objectives, contributing to a greener and more resilient industry.

5 - Forecast Linear Augmented Projection (FLAP): A free lunch to reduce forecast error variance

George Athanasopoulos, Monash University, Caulfield East, Australia, Fin Yang, Rob Hyndman, Anastasios Panagiotelis

A novel forecast linear augmented projection (FLAP) method is introduced. FLAP provably reduces the forecast error variance of any unbiased multivariate forecast without introducing bias. The method first constructs new series as linear combinations of the original series. Forecasts are then generated for both the original and new series. Finally, the full vector of forecasts is projected onto a linear subspace where the constraints implied by the combination weights hold. These new series are called components. It is proven that the trace of the forecast error variance is non-increasing with the number of components, and mild conditions are established for which it is strictly decreasing. It is also shown that the proposed method achieves maximum forecast error variance reduction among linear projections. The theoretical results are validated through simulations and two empirical applications based on Australian tourism and FRED-MD data. Notably, using FLAP with Principal Component Analysis (PCA) to construct the new series leads to substantial forecast error variance reduction.

TB05

Summit - 324

Exploring Equilibrium in Games

Contributed Session

Chair: Paul Messinger, University of Alberta School of Business, Edmonton, AB, Canada

1 - Computing Extended Proper Equilibria and Perfect d-Extended Proper Equilibria with Compact Formulations

Chuangyin Dang, City University of Hong Kong, Kowloon, Hong Kong, Yiyin Cao

The concept of extended proper equilibrium was formulated by Milgrom and Mollner (2021) as a strict refinement of proper equilibrium, which further alleviates the ambiguity of Nash equilibrium. Nonetheless, the formulation of extended proper equilibrium has an unnecessarily strong requirement on strategy perturbations and consequently, demands extremely high computational precision so that it is costly to compute such an equilibrium when the problem size is even moderate. To mitigate this difficulty, we propose the concept of perfect d-extended proper equilibrium, where d is a given number for controlling the degree of extended properness. For each given game, there exists d_0 such that every perfect d-extended proper equilibrium is an extended proper equilibrium provided that d <= d_0. Examples show that a perfect d-extended proper equilibrium remains an extended proper equilibrium even when d is very large. Differentiable path-following methods are developed to compute extended proper equilibria and perfect d-extended proper equilibria. Incorporating logarithmic-barrier terms into payoff functions with an extra variable, we constitute a perturbed logarithmic-barrier game in which all the players jointly solve against a given mixed strategy profile a strictly convex optimization problem. We establish the existence of smooth paths to an extended proper equilibrium and a perfect d-extended proper equilibrium. To reduce the computational burden, we exploit the sorting networks to derive compact formulations of extended proper equilibrium and perfect d-extended proper equilibrium, which consist of a polynomial number of variables and equations. Comprehensive numerical results are reported to show the effectiveness and efficiency of the methods.

2 - Further Extended Proper Equilibrium: Theory, Computation, and Applications

Yiyin Cao, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Chuangyin Dang

Extended proper equilibrium, a refinement of Myerson's (1978) proper equilibrium, was initially formulated by Milgrom and Mollner (2021). In our research, we aim to refine the concept of extended proper equilibrium by introducing the notion of further extended proper equilibrium. This is achieved through the construction of a perturbed strategy space with the property that each of its linear inequality constraints with upper bounds consists of at least one strategy of every player. We show that every further extended proper equilibrium is an extended proper equilibrium. To address the challenges associated with numerical precision in further extended proper equilibrium, we propose the concept of perfect d-further extended proper equilibrium, where the parameter d (0 < d < 1) controls the degree of further extended properness. We rigorously prove that, for sufficiently small values of d, every perfect d-further extended proper equilibrium is indeed a further extended proper equilibrium. Moreover, we exploit logarithmic-barrier games to establish the existence of a smooth path to a further extended proper equilibrium and a smooth path to a perfect d-further extended proper equilibrium, which enable the development of differentiable path-following methods for computing such equilibria. Numerical results further confirm the effectiveness and efficiency of the methods.

3 - Analysis and Discussion of The WGA Strike of 2023

Lara Inci Kullu, Trinity Preparatory School, Orlando, FL, United States

In this paper, we study the Writers Guild of America strike of 2023. Writers are critical suppliers of the entertainment industry, which is a major service industry in the U.S. The strike affected not just the movie and television studios, but also the whole economies of California, Georgia, and New Mexico, placing the economic cost at more than 6 billion dollars. We look at the problem with analytical tools from labor economics and discuss equitable solutions for the industry as a whole and future research directions.

4 - Search and Predict: A Geometric-Inspired Algorithm for Finding Nash Equilibria

Taemin Kim, Rensselaer Polytechnic Institute, Troy, NY, United States, James Bailey

A standard approach for finding Nash equilibria in multiagent games is to apply online optimization algorithms such as gradient descent or multiplicative weights where the time-average of agent strategies are known to converge to the set of Nash equilibria. In this stream of research, we prove geometric properties for gradient descent in zero-sum and coordination games. We then exploit these geometric properties to directly compute a Nash equilibrium after accessing the output of a constant number of iterations of gradient descent. The proposed methods have the potential to drastically improve learning rates while simultaneously decreasing the number of queries/training points needed to find equilibria.

5 - Equilibrium Refinement for Asymmetrical Stackelberg Leadership Duopolies

Paul Messinger, University of Alberta, Edmonton, AB, Canada, Arjang Nikbakht, Bora Kolfal

Often in Stackelberg duopoly both firms prefer the same role (e.g., leadership). The current paper proposes an equilibrium refinement for role selection in such cases. The refinement shows how applying a performance criterion assures a form of perceptual congruence between the equilibrium outcomes and the assumed player roles: a very dominant (weak) firm cannot consistently be assumed to be the follower (leader). This paper gives a general tool for identifying and ruling out implausible leadership structures for the purpose of model design. Also, ruling out implausible equilibria makes the equilibrium predictions more managerially relevant. Lastly, applying this approach leads to game predictions that can be tested empirically.

Our approach assumes the difference in two focal parameters, like production costs, is the only source of asymmetry in the duopoly. Second, we identify all cases where at least one role selection is feasible based on both firms being profitable at the equilibrium. Third, we partition the feasible cases in the focal parameter space based on whether they satisfy particular performance criteria. This partition identifies qualitatively different regions that are characterized by various properties that we discuss. Finally, we categorize the implications for role selection for each region. We show six potential regions emerge for a general model; then we extend the model to include additional conditions that go beyond profitability of firms, which gives rise to three new regions.

TB06

Summit - 325

Optimization and Game Theory in Power Systems and Sports

Contributed Session

Chair: Sreerag Choorikkat, National Tsing-Hua University, Hsinchu, N/A, Taiwan

1 - Dual Conic Proxies for AC Optimal Power Flow

Guancheng Qiu, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Tanneau, Pascal Van Hentenryck

In recent years, there has been significant interest in the development of machine learning-based optimization proxies for AC Optimal Power Flow (AC-OPF). Although significant progress has been achieved in predicting high-quality primal solutions, no existing learning-based approach can provide valid dual bounds for AC-OPF. This paper addresses this gap by training optimization proxies for a convex relaxation of AC-OPF. Namely, the paper considers a second-order cone (SOC) relaxation of AC-OPF, and proposes a novel architecture that embeds a fast, differentiable (dual) feasibility recovery, thus providing valid dual bounds. The paper combines this new architecture with a self-supervised learning scheme, which alleviates the need for costly training data generation. Extensive numerical experiments on medium- and large-scale power grids demonstrate the efficiency and scalability of the proposed methodology.

2 - The Madness of Multiple Entries in *March Madness*

Jeff Sylvestre-Decary, University of Connecticut, Storrs, CT, United States, David Bergman, Carlos Cardonha, Andrea Lodi, Jason Imbrogno

This paper explores multi-entry strategies for betting pools related to single-elimination tournaments. In such betting pools, participants select winners of games, and their respective score is a weighted sum of the number of correct selections. Most betting pools have a top-heavy payoff structure, so the paper focuses on strategies that maximize the expected score of the best-performing entry. There is no known closed-formula expression for the estimation of this metric, so the paper investigates the challenges associated with the estimation and the optimization of multi-entry solutions. We present an exact dynamic programming approach for calculating the maximum expected score of any given fixed solution, which is exponential in the number of entries. We explore the structural properties of the problem to develop several solution techniques. In particular, by extracting insights from the solutions produced by one of our algorithms, we design a simple yet effective problem-specific heuristic that was the best-performing technique in our experiments, which were based on real-world data extracted from recent *March Madness* tournaments. In particular, our results show that the best 100-entry solution identified by our heuristic had a 2.2% likelihood of winning a $1 million prize in a real-world betting pool.

3 - Strategic Interactions in the EV Li-Ion Battery Industry: a Game Theoretical Analysis of the Equilibrium Between Li-Ion Manufacturers, Recycling Companies, and Lithium Suppliers in An Epec Context

SREERAG CHOORIKKAT, National Tsing-Hua University, Hsinchu, Taiwan, YU- CHING LEE, HSIN- WEI HSU

This study delves into the strategic dynamics of the lithium-ion (Li-ion) battery supply chain using a game-theoretical approach and an Equilibrium Problem with Equilibrium Constraints (EPEC). It addresses the interactions among key players: Li-ion battery manufacturers, recycling companies, and lithium suppliers. Manufacturers and recyclers act as leaders in a competitive market, with lithium suppliers serving as followers, aiming to minimize costs under quantity, taxation, and budget constraints. A significant focus of this research is the integration of environmental taxes and subsidies. These economic instruments are explored for their potential to drive sustainable practices by influencing the production and recycling behaviors within the industry. The mathematical modeling of the study confirms the existence of equilibrium solutions that strategically utilize environmental taxes to optimize the sustainability of industry practices. Crossover points, or critical tax levels that lead to shifts in manufacturing and recycling behaviors are a crucial aspect of this study. These points are analytically identified to show how changes in tax incentives can lead to more environmentally friendly industry practices. The study's findings indicate that strategic use of crossover points in tax policies can significantly influence the quantities of Li-ion batteries manufactured and recycled, pointing to the potential of fiscal policies to foster a more sustainable battery supply chain. This research contributes valuable insights into policymaking by comprehensively understanding how economic incentives can be aligned with environmental goals.

4 - Enhanced critical node detection with beam search: A heuristic-agnostic approach

Faraz Khoshbakhtian, University of Toronto, Toronto, ON, Canada, Eldan Cohen, Dionne Aleman

This study presents a novel integration of beam search for the critical node detection problem (CNDP), a well-known NP-Hard combinatorial optimization problem with applications ranging from healthcare to social network influence. Beam search is a widely employed heuristic search algorithm for generating high-quality sequences of tokens in combinatorial search spaces. We leverage the versatility of beam search and tailor it for CNDP, setting our approach apart from existing heuristics by its heuristic-agnostic nature. This versatility enables compatibility and seamless application with various CNDP scoring heuristics. We introduce the doubly scored beam search for CNDP that ensures solutions are optimized both with respect to model outputs and the ground-truth connectivity objective. Our beam search-enhanced CNDP pipeline offers priority interpretation and an ordering of critical nodes, therefore, producing solutions that are effective at every budget level, not just at a threshold. Our doubly scored beam search algorithm has a time complexity of $O(b*w*n \log n)$ in sparse networks of size n, beam width w and budget b, which enables efficient exploration of the solution space, maintaining a balance between exploration and exploitation. We provide the first systematic evaluation of beam search approaches for CNDP in the literature, providing results for singly and doubly scored beam search, standard, diverse, and stochastic beam search, and classical and learning-based node criticality scoring mechanisms. Our empirical evaluations across diverse node-scoring heuristics, beam search variations, and network topologies demonstrate that our doubly scored beam search improves CNDP heuristics, often by double digit margins.

5 - A Mathematical Model for Break Minimization in Round Robin Tournaments

Nasser Salmasi, SimpleRose, Wilmington, NC, United States, Ali Hassanzadeh

In this study, we address the break minimization problem in round-robin sports scheduling tournaments. We focus on a mixed-integer linear programming (MILP) model designed to minimize consecutive home or away games, which are termed 'breaks'. Our framework incorporates advanced constraints and valid inequalities that significantly enhance computational efficiency. This allows for the rapid generation of schedules with minimal breaks for tournaments involving up to 30 teams. We present a detailed analysis of the model's complexity, demonstrating that our approach can solve instances previously deemed challenging by existing mathematical frameworks in mere minutes.

TB07

Summit - 327

Optimization Approaches for Machine Learning

Contributed Session

Chair: Robert Parker, Los Alamos National Laboratory, Los Alamos, NM, United States

1 - Improving the efficiency of Logic-Based Benders Decomposition for large-scale optimization

Paulo Nascimento, Faculty of Sciences and Technology of the University of Coimbra, Coimbra, Portugal, Cristóvão Silva, Samuel Moniz, Christos Maravelias

Large-scale optimization, especially in NP-hard problems, presents significant challenges. To find near-optimal solutions for such problems within a reasonable time frame, (meta-)heuristics are commonly employed. However, when a proven optimal solution is needed, decomposition methods may be the only practical option. Logic-based Benders decomposition (LBBD) expands on traditional Benders decomposition by accommodating any type of optimization or constraint satisfaction problem as sub-problems, making it suitable for a wide range of complex optimization challenges. The problem is divided into a master problem and one or more sub-problems, which exchange information to converge to the optimal solution. Like most decomposition methods and optimization tools, however, the effectiveness of LBBD is heavily dependent on its construction. This study investigates methods to enhance LBBD's efficiency in the context of nesting and scheduling problems in Additive Manufacturing, and explores how these improvements can be generalized to other problems. Various techniques were examined to boost the efficiency of the models used in both the master and sub-problems. Moreover, extracting maximum information from the sub-problems can be computationally expensive, but this is usually offset by generating more effective cuts in the master problem. Additionally, when symmetrical optimal solutions in the master problem cannot be avoided, prioritizing those that are consistent with the solutions from the sub-problems can significantly accelerate the process. The findings suggest that strategic design of an LBBD can influence its computational performance by orders of magnitude, allowing to obtain optimal solutions in a short period of time, ultimately contributing to faster and better decision-making.

2 - GPU accelerated Linear Programming Solver for large problems

Nicolas BLIN, NVIDIA, Paris, France, Katrine Scheel, Piotr Sielski, Akif Coerduek, Rajesh Gandham, Thejaswi Nanditale, Alex Fender, Joe Eaton

Linear programming (LP) is an essential tool for optimization and has applications in resource allocation, production planning, logistics and, as a backbone for Mixed Integer Programming (MIP) solvers. Most of the research and development has been focused on CPUs and direct solvers which require matrix factorization, thus limiting the size of the problems we can solve in a reasonable time. Recent research has been more focused on First Order Methods (FOM) which are more well-suited for parallelization on the latest computing architectures. We introduce a GPU implementation of PDLP, a new FOM algorithm proposed in [1]. It allows us to scale to Linear Programs containing millions of variables and constraints. We observe up to 100x speed up compared to the CPU-based solver.

In this talk, we will discuss how using state-of-the-art CUDA libraries and programming tools allowed us to map PDLP fully and in a highly optimized manner onto the GPU. Furthermore, we will showcase how fine-tuning the hyper-parameters using genetic algorithms allowed us to further gain significant performance improvement. We will conclude by presenting scalability and performance results.

[1] Practical Large-Scale Linear Programming using Primal-Dual Hybrid Gradient

3 - Convergence Reliability of Surrogate and Implicit Formulations for Nonlinear Optimization

Robert Parker, Los Alamos National Laboratory, Los Alamos, NM, United States, Sergio Bugosen

Nonlinear optimization is a powerful tool for engineering and operations research practitioners. While second-order interior point methods such as IPOPT and KNITRO are popular, their successful convergence can be sensitive to model parameters, initial guess, and scaling factors. When this sensitivity is caused by a subsystem of equality constraints that is square and nonsingular (i.e. can be simulated in isolation), several methods are available to improve convergence reliability by replacing these constraints with “easier” functions or constraints. Here, we compare three alternatives that have been proposed to replace such constraints: Sparse polynomial surrogates computed by ALAMO [1], smooth neural network surrogates [2], and implicit functions [3]. In a recent chemical process optimization case study [4], an ALAMO surrogate leads to the most reliable convergence, followed by an implicit function formulation, followed by a smooth neural network surrogate, all of which outperform the original, full-space model.

[1] Alison Cozad, Nikolaos V. Sahinidis, and David C. Miller. Learning surrogate models for simulation-based optimization. AIChE Journal, 2014.

[2] Francesco Ceccon, Jordan Jalving, Joshua Haddad, Alexander Thebelt, Calvin Tsay, Carl D Laird, and Ruth Misener. Omlt: Optimization & machine learning toolkit. Journal of Machine Learning Research, 2022.

[3] Robert Parker, Bethany Nicholson, John Siirola, Carl Laird, and Lorenz Biegler. An implicit function formulation for optimization of discretized index-1 differential algebraic systems. Computers & Chemical Engineering, 2022.

[4] Sergio I. Bugosen, Carl D. Laird, and Robert B. Parker. Process flowsheet optimization with surrogate and implicit formulations of a Gibbs reactor. *Accepted to FOCAPD 2024*.

4 - Large Language Models to Improve Defense Acquisition

Maurice Hudson, University of Alabama at Huntsville, Huntsville, AL, United States, Hieu Pham, Yi Tan

Many organizations are faced with the growing challenge of generating, analyzing, and tracking enormous amounts of text-based data to meet organizational needs. Simultaneously, these organizations are facing high rates of retirement which can lead to institutional knowledge loss and an inability to quickly recruit and sustainably retain personnel with in-demand expertise and knowledge. These challenges are particularly salient in the defense industry due to the complex and bureaucratic nature of Department of Defense (DoD) acquisition and the long lifecycle of those systems. This creates an opportunity to utilize LLM to improve the efficiency of businesses process. This study investigates the development of LLM Application to develop responses to DoD Request for Proposals (RFPs) with the goal of automating portions of the proposal development process and reducing cycle time.

5 - Eligibility threshold selection in social assistance programs: Targeting the poor from self-targeted population

Yuanzheng Ma, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Guodong Lyu, Chung Piaw Teo, Huan Zheng

Social assistance programs (SAPs) aim to utilize limited budget to serve as many households in need as possible. Targeting is the process of identifying and reaching out to households truly in need. However, targeting is challenging due to 1) limited prediction power on households’ welfare levels, and 2) the applicants of SAPs are biased samples of the whole population due to self-targeting effect. These factors can result in high exclusion errors and leaving budget underutilized in SAPs. To improve targeting performance, we focus on the eligibility threshold decision, which determines applicants’ eligibility based on their predicted welfare levels. We first characterize the self-targeted applicant population in distribution shift and population scale, constructing corresponding estimators with performance guarantees; Then we provide a new policy embedding the estimators into eligibility threshold selections. The numerical results show that the proposed policy significantly outperforms the benchmark policy from practice and covers more people truly in need, especially when welfare prediction power is limited. Beyond SAPs, self-selection and threshold-based control decisions are widespread in many other areas, where we believe the proposed method can also provide potential benefits.

TB08

Summit - 328

Innovative AI and Strategic Decisions: From NFTs to Enterprise Security

Contributed Session

Chair: Rohan Korde, Arizona State University, Tempe, AZ, United States

1 - Propensity for Job Replacement by Artificial Intelligence: An Empirical Analysis

Shenghan Xu, University of Idaho, Moscow, ID, United States, Eric Stuen

The impending integration of Artificial General Intelligence (AGI) presents a monumental shift in the global economy, prompting a critical examination of its impact on industries and career trajectories. This research aims to address the pressing business problem of anticipating and navigating the disruptive effects of AGI on employment landscapes. Drawing on a dataset of 4,762 occupations and employing a mathematical model to quantify job replacement probabilities, the study identifies key findings regarding various industries' susceptibility and educational levels' susceptibility to AGI disruption. It reveals an inverse relationship between education level and replacement probability, with doctoral or professional degrees offering greater resilience. Additionally, the paper explores emerging careers in sectors less susceptible to automation, such as the "Pleasure Economy." This study contributes to existing knowledge by providing insights and guidance to individuals, industries, and policymakers, enabling informed decision-making and adaptation strategies in the face of AGI-driven transformations. It emphasizes the importance of higher education in mitigating the impacts of technological displacement and highlights potential avenues for future job opportunities. Methodologically, the research employs rigorous analysis of contemporary economic theories coupled with examining private investment trends and emerging career trajectories. The dataset allows for a comprehensive assessment of job replacement probabilities across industries and educational levels. The main results underscore the significant variations in replacement probabilities across industries and educational levels, with implications for future workforce planning and educational frameworks. The findings emphasize the need for proactive strategies to foster an equitable and prosperous future amidst the evolving AI-driven landscape.

2 - Zero-Trust in LLM-integrated applications in enterprise networks

Deepti Singh, California State University - Long Beach, Long Beach, CA, United States

The integration of large language models (LLMs) into enterprise applications introduces new cybersecurity risks and challenges to traditional security models. We apply Zero Trust principles to develop a comprehensive access-controlled security framework that addresses the unique security challenges posed by LLM-integrated applications within corporate environments.Unlike a traditional enterprise application, an LLM-integrated application can work in dual-modality - user and process. As a user it can interact with the system, input data, and request actions, much like a human user would. An LLM-integrated application can monitor network logs, access corporate resources, and generate reports. As a process, it can handle requests, generate responses, and manage data in a traditional secured network infrastructure. It can respond to queries, give recommendations, or automate tasks. This dual functionality disrupts the traditional distinction between users and processes. Although it presents new opportunities for enhanced functionality and user interaction, it introduces new challenges for system design and security. Our framework involves validating, monitoring, and enforcing strict access control policies for LLM-integrated applications, with a focus on preventing prompt injection, data poisoning, and unauthorized access. The proposed ZT-AI-Access framework enhances cybersecurity by ensuring robust data protection and compliance with corporate policies, mitigating the risks associated with the unmanaged adoption of AI technologies which are not necessarily confined the an enterprise-owned network boundary.

3 - Multiple orders per job (MOJ) scheduling problems for orders with due dates and release times in a two-stage permutation flowshop

Rohan Korde, Arizona State University, Tempe, AZ, United States, John Fowler, Lars Moench

Amazon, Alphabet, Microsoft, OpenAI, and other firms are driving the innovation in large language models (LLMs), generative pre-trained transformer (GPT) applications, artificial intelligence (AI) applications, and cloud computing. This has created a significant demand for tensor processing unit (TPU), application-specific integrated circuit (ASIC), field-programmable gate array (FPGA), graphics processing unit (GPU), and other semiconductor chips. The demand for semiconductor chips is expected to keep rising in the months and years ahead, as increasing number of companies adopt AI models to make smarter decisions. In a 300-mm wafer fabrication facility semiconductor chips are manufactured in reentrant process flows using hundreds of tools. Front opening unified pods (known as FOUPs or jobs) transport 300-mm silicon wafers between the tools in the fab using an automated overhead hoist transportation system (OHT). The number of FOUPs are typically far fewer than the number of customer orders received. In some cases, customer orders are prioritized and can be combined into a single FOUP depending on certain factors. The FOUPs are then scheduled to be processed on tools to minimize a given performance measure. These scheduling problems are known as "multiple orders per job" (MOJ) problems. We study MOJ scheduling problems to minimize maximum lateness, number of tardy orders, weighted number of tardy orders, total tardiness, and total weighted tardiness for orders with equal and unequal release times in a two-stage permutation flowshop. We design full-factorial experiments to check optimality (small-sized instances) and solution quality (large-sized instances).

4 - Why Firms Buy: Dynamic Games of Acquisition Among Tech Giants

Boyoon Chang, University of Oregon, Eugene, OR, United States

Information and technology markets often feature high market concentration and frequent acquisitions. Understanding the motives underlying these acquisition strategies is therefore a first-order question for policy makers and practitioners. Using a dynamic model, I analyze acquisition records of the “big five” (or “tech giants”) – Google, Amazon, Microsoft, Meta, and Apple. I characterize the firms’ acquisition decisions observed in the data as a Markov perfect equilibrium that is driven by their scale economies and defensive motive. I find that the motive to keep up with their rivals can explain a major share of their acquisition behaviors, sometimes overshadowing their internal motives of scale economies. I show that a company might opt to acquire not necessarily because it can benefit from technological synergies generated from accumulating similar targets, but to align with the trajectories of their rivals to mitigate the risk of falling behind.

5 - Prediction of Violence Victimization Among Children and Adolescents Using Machine Learning

Jennifer Mendoza-Alonzo, Division of Injury Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States, Ketra Rice, Elizabeth Parker, Keming Yuan

Violence victimization among young people is a significant public health problem in the United States. In 2021, there were over 3,000 homicides and over 160,000 emergency department visits for nonfatal assault injuries among young people ages 10-19 years. Violence victimization has significant long-term effects on the physical, social, and emotional health of youth, families, and communities. This study aims to predict violence victimization using social-ecological factors across the individual (i.e., race and ethnicity, attention functioning, gender identity), relationship (i.e., household financial stress, family relationship, peer relationships), and community (i.e., neighborhood safety) levels of the social ecology. Using 2019-2022 data from the Adolescent Brain Cognitive Development® (ABCD) study, we implemented classification models to understand this relationship using different machine learning algorithms such as logistic regression, support vector machines and eXtreme Gradient Boosting. Preliminary results show statistically significant associations between violence victimization and attention functioning ($p < 0.001$), family relationship ($p<0.001$), and neighborhood safety ($p=0.001$). Low attention functioning was found to have a 1.2 greater odds of victimization compared to adolescents with high attention functioning and living in a safe neighborhood was found to have a 0.60 lower odds of victimization compared to adolescents living in unsafe neighborhoods. The outcomes of the study revealed significant insights into the risk of violence victimization among adolescents and highlighted the importance of leveraging advanced computational techniques in addressing complex issues. Continuing research in this domain holds promise for further refining predictive models and enhancing their applicability in real-world settings to safeguard the well-being of adolescent populations.

TB09

Summit - 329

Operating Online Platforms in the New Age

Invited Session

Auctions and Market Design

Chair: Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A

1 - Content Rotation in the Presence of Satiation Effects

Xiao Lei, University of Hong Kong, Hong Kong, Hong Kong, Beichen Wan, Shixin Wang

Service providers often employ rotational schedules for offering products such as fitness courses, game content in online video games, and new episodes on content-sharing platforms. These rotations not only reduce the operating costs associated with constant offerings, but also moderate consumer satiation caused by frequent consumption of the same product. Although simple designs are prevalent in practice, there is a lack of guidance on designing effective rotational schedules. In this paper, we explore the design of rotational schedules for service providers to maximize long-run average profits. We develop an analytical model based on utility theory, which connects consumer consumption with satiation and product correlation. Our investigation covers various scenarios, including single product offerings, regular and rotational product pairs, and multiple product assortments. For a single product, we characterize the optimal frequency. For a regular and rotational product pair, we observe an interesting ripple effect in consumer dynamics. In this case, we examine the optimal frequency and selection of the rotational product, which depend on reciprocal influences between the two products. In scenarios involving multiple products, we demonstrate that as the number of products increases, the assortment depth converges to a constant, and the cycle length should grow linearly. Additionally, we find that product permutations have a marginal impact on total consumption. To address general situations, we introduce a numerical method based on mixed-integer programming and conduct numerical studies using both synthetic and real data. Our findings and insights offer a first step towards understanding the design of content rotation.

2 - Dynamic Persuasion Strategies for Mitigating the Spread of Fake Content

Wenjuan Li, HKUST Business School, Clear Water Bay, Hong Kong, Dongwook Shin

Digital platforms allow users to create and share content on an unprecedented scale, but fake news and misinformation can quickly spread, posing a threat to society. Our research investigates how platforms can detect fake content and influence user actions using personalized recommendation messages. Our model involves random inspections of content by the platform, allowing it to determine the authenticity of the content once it has been inspected. We characterize optimal inspection and recommendation policies, finding that ad-based platforms may recommend fake content, while subscription-based platforms reveal true fakeness. We compare these results with public recommendation policies. Furthermore, we investigate the impact of the inspection and recommendation policies on content creators.

3 - Contract Design in Fulfillment by Platform

Xingyu Lin, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Jingqi Wang, Jiaqi Lu, Yilun Chen

We study a third-party seller selling a product through an online platform, where the seller can fulfill his demands either by inventory stored in the platform's fulfillment center or by in-house inventory. We characterize the optimal FBP (Fulfillment by Platform) contract with sales profits sharing and inventory storage fee offered by the platform. We show that FBP coordinates the supply chain with a single seller. With the optimal contract, the platform extracts all surplus above the seller's reservation profit and will give a subsidy to the seller. For further discussion, we find that FBP also coordinates the supply chain with multiple sellers selling substitute products. We also analyze how FBP influence the platform's profit compared with FBS (Fulfillment by Seller), and demonstrate that this impact is related to seller's type (production costs, in-home inventory costs, etc). This leads us to a further model extension of the contract design when sellers have private information.

4 - Optimal Signaling of Supply Competition in Decentralized Platforms

Yanwei Sun, Imperial College Business School, London, United Kingdom, Niloofar Zamani Foroushani, Zhe Liu, Jiahua Wu

We consider a specific type of decentralized platforms prevalent in the service sector, characterized by a system where customers can freely post jobs or tasks on one side, while service providers must pay a fee to unlock the ability to communicate with customers. The purchase of job leads is not exclusive, permitting several service providers to access the same job lead. We employ a Bayesian persuasion framework to explore how the platform can leverage its informational edge concerning the supplier base to strategically influence service providers' decisions, with the goal of maximizing its revenue. We characterize the structure of the optimal signaling mechanism, wherein the platform recommends a deterministic number of service providers to participate when the pool size is smaller than a threshold; otherwise, the recommendation randomizes between two consecutive numbers. Then, we endogenize the platform's pricing decision and show that coupled with an optimal fixed price, the optimal signaling mechanism generates a higher expected revenue than the optimal state-dependent pricing mechanism. Besides, more notably, within the policy space where information design and state-dependent pricing are considered jointly, information design with one fixed price remains optimal.

TB10

Summit - 330

Opioid Crisis and Health Systems Management

Contributed Session

Chair: Yijie Chen, Centers for Disease Control and Prevention, Atlanta, GA, United States

1 - Investigating Factors Associated with Willingness to Receive Treatment Among Adults Reporting Fentanyl use: a Machine-Learning Approach for Feature Extraction and Interpretation

Yijie Chen, Centers for Disease Control and Prevention, Atlanta, GA, United States, Nisha Nataraj, Xinyi Jiang, Matthew Gladden, Gery Guy, Deborah Dowell

The opioid overdose crisis has rapidly evolved with the availability of illegally made fentanyl (IMF), reshaping patterns of substance use and treatment engagement. In 2022, approximately 90% of the estimated 81,806 opioid-overdose deaths involved synthetic opioids excluding methadone, primarily IMF. Recent reports indicate that only about 1-in-3 individuals with an opioid use disorder received any past-year substance use treatment. This study investigates factors associated with willingness to receive treatment among individuals who use IMF. The National Addictions Vigilance Intervention and Prevention Program Addiction Severity Index-Multimedia Version tool collects a convenience sample from adults aged ≥18 years assessed for substance use treatment. Leveraging a natural gradient boosting (NGBoost)-based algorithm, we analyzed treatment assessment data (2018-2023) from adults who reported past 30-day IMF use, with variables including demographic characteristics, routes of administration, and concurrent substance use. We also incorporated spatial-temporal dynamics analysis to account for regional differences in fentanyl availability and use patterns across the United States. Self-reported days experiencing drug-related problems in the past 30 days, severity of family problems (index reflecting frequency and conflicts), and severity of psychiatric problems (index reflecting frequency and distress), were the top three variables associated with willingness to receive treatment. A higher willingness was associated with more frequent drug use, higher frequency and intensity of family problems, and experience of psychiatric symptoms, with complex interaction effects. Efforts to connect people to evidence-based treatment, before exacerbation of family conflicts or psychiatric issues, is important, especially given that IMF use is associated with increased fatal overdose risk.

2 - The Complexity-Interpretability Tradeoff Between Risk-Scoring Methods: Implications for State Medicaid Agencies

Shailesh Divey, University of Oregon, Eugene, OR, United States, Nickolas Freeman, Dwight Lewis, Jason Parton

State Medicaid agencies rely on risk-scoring methods to project costs, budget, and identify high-cost categories that can be targeted for preventative interventions. Currently, many risk-scoring approaches exist and vary widely on the dimensions of sophistication/complexity and interpretability. For example, the popular CDPS risk-scoring method is open-source and utilizes a straightforward linear regression method to generate coefficients that can be inspected to see the impact of patient-related variables on the assigned risk score. On the other end of the spectrum, the proprietary MARA model has been shown to offer more accurate risk-score estimates, however, it offers little with respect to interpretability. Using data from Alabama Medicaid collected over three years (2016–2018), this research investigates options for risk-scoring that span the spectrum of sophistication and interpretability. Insights from this research can inform decision-makers regarding the tradeoffs that arise among these two dimensions. We conduct experimentation on feature choices that provide insight regarding the impacts of feature engineering and feature selection on resulting models. Ultimately, the insights gained from Medicaid risk-scoring models can be leveraged to allocate resources effectively and improve healthcare access for underserved communities.

3 - Assessing the Equitable Distribution of Naloxone Locations

Melika Jahan Beikloo, Clemson University, Clemson, SC, United States, Mary Beth Kurz, Bryan Lee Miller, Thomas Sharkey, Chelsea Spence

The opioid epidemic remains a critical public health concern, with South Carolina (SC) experiencing its share of fatal consequences, including high rates of opioid overdose deaths. In this study, we employ Geographic Information Systems (GIS) and regression analysis to investigate the equitable access to naloxone, a crucial opioid overdose reversal medication, focusing on the most affected counties in SC.

Utilizing GIS, we spatially analyze the distribution of naloxone resources relative to opioid overdose hotspots, identifying potential disparities in access. Additionally, we employ regression modeling to assess the relationship between naloxone distribution and opioid overdose death rates. By incorporating variables such as income, demographics, population density, and healthcare access, our regression analysis seeks to explore the drivers of naloxone access and overdose outcomes.

This study aims to provide insights into the equity of naloxone distribution in addressing the opioid crisis, particularly in areas with the greatest need. Findings from this research can inform evidence-based interventions and policy initiatives to promote equitable access to naloxone and reduce opioid-related harms in SC communities.

4 - Balancing Enforcement and Access in Prescription Drug Networks via Multi-Objective Optimization

Monika Tomar, Purdue University, West Lafayette, IN, United States, Abhishek Ray, Mario Ventresca

The burgeoning crisis of prescription drug misuse, particularly opioids, poses complex challenges that intersect public health, criminology, and network science. Prescription drug networks operate within a dual paradigm, serving both legitimate medical needs and facilitating illicit drug diversion. This duality necessitates a nuanced approach to intervention, aiming to dismantle harmful activities without disrupting the legitimate supply chain essential for patient care. Drawing upon insights from deviant behavior into drug diversion dynamics and practices for harm reduction, this paper examines the intricate balance between enforcing drug misuse prevention and ensuring uninterrupted access to necessary medications. We propose a multi-objective combinatorial decision-making framework that prioritizes node removal by targeting key nodes whose removal would maximally disrupt illicit activities with minimal impact on legal drug distribution and patient access. The paper offers a detailed formulation of a combinatorial decision problem keeping in mind the interaction of intervention actions over a network of suppliers, providers/pharmacies and patients. The paper further explores the intricate tradeoffs between the various objectives by analyzing the various pareto fronts obtained by solving the problem using evolutionary algorithms.

5 - Smarter Decisions about Drug Overdoses for a Better US: Usefulness-Focused Data Profiles and Risk Modeling from the Southern Regional Drug Data Research Center

Joshua Eyer, University of Alabama, Tuscaloosa, AL, United States, Amanda Cox

The overdose crisis is experiencing its fourth wave, characterized by a spike in deaths from opioids combined with illicit stimulants. Decision makers across sectors experience challenges in responding to this on-going crisis despite limited access to comprehensive data.

The Southern Regional Drug Data Research Center (DDRC) represents an innovative initiative funded by the Bureau of Justice Assistance to promote knowledge across affected parties in a 17-state region of the south US. Housed in the Institute of Data and Analytics in the Culverhouse College of Business at the University of Alabama, the DDRC features a multidisciplinary team who will present information about the aims and method of the project to develop a large multisource (public + proprietary) data repository and to promote the development, accessibility, and transmission of knowledge from the repository to strategic stakeholders. These include public health, public safety, community organizations, nonprofits, journalists, students, and academic researchers.

The project integrates high-value public datasets (e.g., CDC, NIBRS, US Census, ARCOS) and facilitates their use for research at various levels. Educational materials and efforts will enhance research communication to promote users' ability to utilize results. DDRC analytic reports will provide innovative insights, bridging state boundaries and synthesizing data from disparate sources. An important aspect of this work and presentation will be efforts to develop substance use risk metrics for the region, complementing county-level profiles generated for our public-facing website and on-demand analytic reports. Attendees will learn how we developed the DDRC, what data we share, and how to use available resources.

6 - Addressing the Challenges of Mining Patient Data from a Procurement Perspective

Stella Hua, Western Washington University, Bellingham, WA, United States

The analysis of big patient data provides enormous healthcare benefits such as increased diagnosis accuracy and improved treatment efficiency. However, the analysis of big patient data also draws criticism from the public and poses challenges for healthcare agencies and IT vendors. This study reviews the procurement processes of several cases involving large scale healthcare data mining. The goal is to identify preventive measures that procurement needs to undertake to help address these criticisms and challenges. These preventive measures add to the dimension of procurement maturity in the procurement literature and provide insights for healthcare providers pursuing data mining services.

TB11

Summit - 331

Design and Optimization for Transit Systems

Contributed Session

Chair: Mehr Sadat Salami, North Carolina State University

1 - Nonparametric Compressed Learning of Dynamical Systems

Boya Hou, University of Illinois,Urbana-Champaign, Urbana, IL, United States

Autonomous systems are just around the corner. Different from ``automated'' agents that operate in a fixed world, ``autonomous'' ones should be able to conceptualize, reason, and act in changing environments without human intervention. Fulfilling such an ambition requires ways to understand the environmental landscape. Under purely data-driven approaches that passively adapt to environments, agents fail to leverage previously learned knowledge and need to solve the problem from scratch every time the environment changes. The mature field of systems theory has enabled the success of model-based decision-making. Yet model identification typically requires fitting parametric models to data from interaction with the environments. In this talk, I will discuss an operator-theoretic approach to learning compressed representations of nonlinear dynamics from data with provable guarantees. We first build an analytically tractable representation of system dynamics via the conditional mean embedding operator that interacts with a reproducing kernel Hilbert space (RKHS). Compared with neural networks, kernel methods are theoretically easier to analyze, easily implementable, and demonstrate excellent empirical performance in data-limited regimes. Still, the scalability often suffers with larger datasets. To counter this difficulty, we allow selective loss in the representation of that operator to control model complexity. I will discuss sample complexity guarantees to learn such operators in centralized and decentralized settings and present applications to reinforcement learning, power system transient stability analysis, and uncertainty propagation.

2 - factors Impacting the interaction of autonomous and legacy vehicles

Roohollah Jahanmahin, Wayne State University, Detroit, MI, United States, Sara Masoud

The coexistence of autonomous vehicles (AVs) and legacy vehicles (LVs) marks a pivotal moment in the evolution of transportation. AVs, empowered by advanced technologies like sensors and artificial intelligence, offer the promise of safer and more efficient roads by reducing human error. However, with legacy vehicles primarily operated by humans, it is imperative to foster harmonious coexistence and devise strategies for their seamless interaction. This necessitates careful consideration of how AVs integrate with LVs and existing infrastructure, addressing regulatory, safety, and ethical concerns. This study delves into the dynamics between AVs and LVs using an immersive driving simulator that merges the DReyeVR project with the Carla simulator. Leveraging virtual reality through HTC Vive Pro Eye, the research emulates a first-person maneuverable ego-vehicle equipped with eye-tracking and immersion enhancements. Extensive datasets capturing real-time eye-tracking metrics and vehicle control inputs enable a convolutional neural network (CNN) model to analyze and forecast driver intention and decision-making behavior. The insights garnered, including crucial gaze patterns and driver focus, deepen our understanding of trust and safety perceptions regarding AVs. Moreover, the study evaluates driver responses such as acceleration, braking, and lane-changing behaviors in close proximity to AVs, pinpointing key factors in driver decision-making processes. These findings aim to refine interaction strategies between LVs and Level 2 and 3 AVs, ultimately fostering the safer integration of AVs into public roadways.

3 - The Impact of Automation on Workers when Workers are Strategic: The Case of Ride-Hailing

Zicheng Wang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Saif Benjaafar, Xiaotang Yang

Motivated by the behavior of drivers on ride-hailing (individual drivers decide whether to work based on the offered wage and where to locate themselves in anticipation of future fares), we examine how the introduction of autonomous vehicles impacts the strategic behavior of human drivers and driver welfare. Specifically, we consider a setting where a ride-hailing platform deploys a mixed fleet of conventional vehicles (CVs) and autonomous vehicles (AVs). The CVs are operated by human drivers who make independent decisions about whether to work for the platform and where to position themselves when they become idle. The AVs are under the control of the platform. The platform decides on the wage it pays the drivers, the size of the AV fleet, and how the AVs are positioned spatially when they are idle. The platform can also make decisions on whether to prioritize the AVs or the CVs in assigning vehicles to customer requests. We characterize the optimal decisions of the platform and contrast those with the optimal decisions in the absence of AVs. We show there are settings where introducing AVs leads to higher wages and more drivers being hired. These results can be explained by the interplay of two counteracting effects: (1) the introduction of AVs provides the platform with an additional source of supply and renders human driver substitutable (*displacement effect*); and (2) having access to and control over AVs enables the platform to influence the strategic behavior of CVs, thereby reducing the inefficiency from self-interested behavior (*incentive effect*).

4 - Applying Autonomous Vehicle Adoption Model in Examining Customers&Rsquo; Adoption Intention of Driverless Cars

Thi Ngoc Anh Tran, Korea University, Seoul, Korea, Republic of, Daeki Kim

The increase in personal vehicle usage prompted by the COVID-19 pandemic emphasizes the potential for a transformative shift within the travel industry through the widespread adoption of self-driving autonomous vehicles (AVs). Previous research has primarily concentrated on creating models for accepting automonous vehicles in developed countries, while overlooking how moderating variables influence these models. Therefore, this research aims to construct and evaluate a theoretical model for the acceptance of autonomous vehicles, highlighting demographic characteristics, individual innovativeness, and driving behaviors as pivotal factors influencing travelers' tendency towards AV utilization. Drawing upon the Autonomous Vehicle Acceptance Model (AVAM) as a foundational framework, this study employs structural equation modeling (SEM) techniques on data gathered from 1008 respondents in Hanoi, Vietnam. Results reveal noticeable differences in the drivers of behavioral intention to embrace AVs between individuals exhibiting high and low levels of personal information technology innovativeness. Notably, the perceived performance expectancy and effort expectancy show a significant impact on attitudes toward autonomous vehicles, subsequently shaping the intention to utilize them. Furthermore, facilitating conditions emerge as a direct influencer on the propensity to adopt self-driving vehicles. The limitations of research method and implications of these findings are discussed, offering insights that may prove valuable to marketers and manufacturers in navigating the landscape of self-driving vehicle adoption.

Key words: Autonomous Vehicle Acceptance Model (AVAM), personal innovativeness, self-driving autonomous vehicle.

5 - Optimal Volunteer Management in Non-profit Settings Balancing Equity and Efficiency

Hiruni Niwunhella, North Carolina State University, Raleigh, NC, United States, Mehr Sadat Salami, Faisal Alkaabneh, Leila Hajibabai

This paper presents a pioneering solution to the vehicle routing problem with pickups and deliveries with equity (VRPPD-e) within the realm of non-profit food distribution. With a focus on multiple volunteers at service centers helping with food deliveries to food-insecure households through food banks, a mixed integer programming model is developed that aims to optimize volunteer assignments, plan cost-efficient routes for their pickups and drop-offs at service centers and back, and ensure equitable service distribution. Considering each volunteer’s time window and vehicle capacity, we aim to maintain a balance among these key factors. We present a scalable solution utilizing a branch-and-price algorithm. This approach transforms the proposed model into a set partitioning problem as a relaxed linear program. The algorithm dynamically refines solutions by iteratively introducing variables. Leveraging the power of branch and price, coupled with column generation, the proposed solution technique systematically enhances the linear programming relaxations. To further optimize the column generation process, we employ a labeling algorithm. This additional step contributes to the effectiveness of our approach, providing a comprehensive strategy for solving complex optimization problems efficiently. This integrated methodology not only redefines the VRPPD-e paradigm but also provides a novel solution to workforce management challenges in the non-profit sector. The preliminary experiments highlight the algorithm’s effectiveness in optimizing volunteer assignments and vehicle routing, emphasizing considerations of fairness and cost under time-window constraints and vehicle capacity restrictions.

TB12

Summit - 332

Methods for Traffic Monitoring

Contributed Session

Chair: Sifa Celik, Eindhoven University of Technology, EINDHOVEN, Netherlands

1 - Vehicle Routing Problems Under Uncertainties: A Comprehensive Survey of Methods and Challenges

Rashid Baimukashev, SDU University, Kaskelen, Kazakhstan, Cemil Turan, Meraryslan Meraliyev, Shirali Kadyrov

This paper provides a systematic survey of Vehicle Routing Problems (VRPs) under uncertainties, focusing on the diverse uncertainties encountered and the methods used to address them. Uncertainties in VRPs arise from several factors, including variable demand, fluctuating travel times, and evolving market conditions. These uncertainties can significantly impact the efficiency and reliability of routing solutions, making effective management critical for logistics and supply chain operations. The survey classifies these uncertainties and reviews the methodologies developed to handle them, spanning exact methods, heuristic approaches, and learning-based techniques. A novel taxonomy categorizing uncertainties and solutions is introduced, aiding in the comprehensive understanding of this complex field. Furthermore, the paper discusses the practical challenges faced by practitioners and decision-makers, emphasizing the gap between theoretical research and its real-world applications. By providing a detailed evaluation of current methods and identifying areas lacking sufficient attention, this survey aims to guide future research directions and enhance the applicability of routing solutions under uncertainty.

2 - Deep Reinforcement Learning for Efficiently Solving the Electric Vehicle Routing Problem with Time Windows and Partial Recharging

Mertcan Daysalilar, University of Miami, Miami, FL, United States, Adam Meyers

Conventional research on the Electric Vehicle Routing Problem (EVRP) often relies on traditional exact optimization methods or metaheuristic approaches, which lack some flexibility to adapt to changing operational conditions and energy demands. In addition, such methods are not well-suited for large-scale routing problems or real-time solutions needed as the modern world attempts to develop nationwide charging infrastructures to support the transition from gasoline-powered vehicles to EVs. Deep reinforcement learning (DRL) has recently been proposed and has been shown to help overcome the above limitations. This study introduces a comprehensive DRL framework to tackle the EVRP with time windows and partial recharging. Implementing time windows ensures that customer demand is met in a predetermined span of time, while implementing partial recharging allows for a more flexible charging policy that can help mitigate the undesirable downtime and operational delays arising from a rigid full recharging policy. The proposed method utilizes an attention model combining a pointer network with a graph embedding layer, which allows for a stochastic policy capable of addressing the complexities of routing EVs. The model, which is trained using a policy gradient with a rollout baseline, is designed to optimize routes and charging strategies for large-scale instances. The primary objective is to evaluate the model's effectiveness in tackling large problem instances that traditional methods cannot efficiently handle. The proposed approach not only contributes to the theoretical advancements in DRL applications for logistics, but also offers practical solutions for real-world EV fleet operations.

3 - Using Connected and Automated Vehicles as Mobile Controllers in Single Lane Roundabouts

Abdullah Farabi, NC State University, Raleigh, NC, United States, Ali Hajbabaie

This study develops a mixed integer non-linear program where connected and automated vehicles (CAV) operate optimally and work as mobile controllers to guide Human Driven vehicles (HVs) through roundabouts. As HVs may incur excess delays due to a lack of safe gaps, CAVs can send accurate indications to HVs for a safe entry to the circular roundabout path. The objective function of this model is to maximize roundabout throughput while controlling the flow of vehicles so that an equitable roundabout operations is achieved. Using Frenet coordinates, the non-linear formulation is simplified to an easier-to-solve mixed integer quadratic program and a model predictive control-based solution technique is developed to optimize CAV trajectories as well as to control HV's gap acceptance behaviors in real-time. The case study results show that the new methodology outperforms conventional traffic operation at roundabouts in terms of various performance measures (e.g., delays, stops, travel times, and throughput). Besides, the proposed model ensures equity between HVs and CAVs and improves roundabout capacity significantly.

4 - Optimal Connected Automated Vehicle Control in Freeway Merge Segments through Distributed Coordination

Fahim Kafashan, North Carolina State University, Raleigh, NC, United States, Ramin Niroumand, Ali Hajbabaie

Complex road segments like freeway merge and diverge segments are critical due to their susceptibility to bottlenecks caused by lane changing maneuvers. This study develops a mixed-integer nonlinear program (MINLP) for optimizing both the longitudinal and lateral movements of connected automated vehicles (CAVs) in these critical segments. This approach uses a hybrid solution technique combining linearization, receding horizon framework, and a vehicle-level decomposition technique to manage the problem's complexity while adapting to dynamic traffic conditions effectively. The methodology is tested in various scenarios to demonstrate its applicability to real-world freeway operations. Demonstrated across various scenarios, this methodology significantly enhances operational efficiency and safety, offering a robust model for future CAV systems integration on freeways. In the merge segment, this model resulted in up to 69.82% reduction in the total travel time and more than 200% increase in average time to collision as an indicator of safety. Similarly, this model improved diverge segments' travel times and average time to collision by up to 24.91% and 250%, respectively.

5 - Efficient Real-Time CAV Trajectory Optimization at Signal-Free Intersections Using a Greedy-Based Heuristic Approach

Ramin Niroumand, Aalto University, Espoo, Finland, Fahim Kafashan, Leila Hajibabai, Ali Hajbabaie, Claudio Roncoli

This study proposes a greedy-based heuristic method to construct connected automated vehicle (CAV) trajectories at signal-head-free intersections that not only does not use a computationally expensive solver but also provides solutions in the order of milliseconds to be used in network-level trajectory optimization frameworks. This methodology constructs multiple trajectories for each CAV and selects the one with the best objective value. In addition, a platooning logic is developed to form platoons consisting of one or more CAVs based on their relative locations and operate platoons of CAVs instead of individual CAVs at a time to achieve lower delay times. The proposed greedy-based solution technique is embedded in a receding horizon framework to further decrease the complexity of the problem and address its dynamic nature. The case study results on an isolated intersection reveal that the proposed methodology is capable of solving the optimization model in under 35 milliseconds under a demand level of 1500 veh/h/lane. The proposed methodology achieves an optimality gap of 4.61% in average delay without relying on commercial solvers.

6 - The Dynamic Delayed Time Window Assignment Vehicle Routing Problem

Sifa Celik, Eindhoven University of Technology, EINDHOVEN, Netherlands, Albert Schrotenboer, Layla Martin, Tom Van Woensel

We explore the critical balance between direct and delayed communication of time windows to customers, a trade-off that significantly impacts customer service, on-time delivery, and transport costs. This paper addresses the Dynamic Delayed Time Window Assignment Vehicle Routing Problem, a complex decision-making challenge faced by service providers managing next-day deliveries. We model our model as a Markov Decision Process (MDP) with a finite time horizon. We propose three policies: Deep Controlled Learning (DCL), the Rollout Algorithm (RO) with stochastic one-step lookahead, and the Multiple Scenario Approach (MSA), modified for our problem setting. DCL, a deep reinforcement learning framework, tackles the computational challenges of large state-action spaces through approximate policy iteration, efficient state sampling, and effective dataset construction. The RO, rooted in simulation, iteratively refines the base policy by approximating the action-value function for state-action pairs, while MSA focuses on likely future scenarios to optimize decision-making. We provide novel insights into optimizing route planning and time window assignments in the context of next-day delivery services. Demand pattern and customer positioning play a crucial role in how the decisions are taken. When customers are dispersed in either in clustered or random-clustered manner across the service area, there is a noticeable increase in their waiting time for time window assignments. There appears to be a relationship between demand variability and waiting time: as demand becomes less consistent, the waiting time for customers increases. This effect is more pronounced when demand has larger mean.

TB13

Summit - 333

Agricultural Supply Chains

Contributed Session

Chair: Harol Gamez Alban, Massachusetts Institute of Technology

1 - Research on dynamic pre-sale pricing decision of fresh agricultural products

yuxiu liang, Southeast University, Nanjing City, Jiangsu Province, China, People's Republic of, Lindu Zhao

Due to the characteristics of long growth cycle, short life cycle and difficult storage of agricultural products, in order to avoid the waste of agricultural products caused by excessive planting by farmers, the pre-sale of agricultural products is worth studying. Farmers obtain order quantity according to pre-planting, so as to obtain market demand in advance, plan planting quantity and time in a targeted way, reduce production uncertainty and risk, and avoid waste caused by over-planting.

This study considers a two-level fresh supply chain system consisting of a single supplier and fresh electricity supplier. Fresh electricity supplier conducts two stages of pre-sale of agricultural products through online channels: Pre-planting pre-sale and pre-mature pre-sale, suppliers and fresh electricity suppliers jointly participate in the current sale period, fresh electricity suppliers reduce the uncertainty of consumers' valuation of agricultural products in the pre-sale period through discount pre-sale, the study considers that in order to promote the enthusiasm of consumers to buy, suppliers and fresh electricity suppliers respectively carry out corresponding value-added services and promotional efforts. Based on the risk of agricultural products planting and the uncertainty of market demand, this paper analyzes the multi-period pre-sale of agricultural products, and studies the decision-making and coordination of supply chain themes under the dynamic pre-sale model of fresh agricultural products.

2 - A Novel Approach to the Continuous Planting and Harvest Scheduling

Maksim Lalic, BioSense Institute, Novi Sad, Serbia, Nikola Obrenovic, Gordan Mimic, Miljana Markovic, Oskar Marko, Sanja Brdar, Vladimir Crnojevic, Ivan Lukovic

Among all food chain losses, post-harvest handling and storage losses are among the most prominent, particularly in less developed countries. Farmers with limited storage and processing capabilities could substantially reduce their yield overflows and consequent unnecessary waste by adopting continuous harvest strategies. Nowadays, the increase in average farmland and the number of crop varieties opens additional room for continuous harvest. However, common planting and harvesting planning practices typically treat each parcel of land in isolation, missing the opportunity of a more integrated approach that would equalize the expected yield during the harvest period, while performing harvest activities in a reduced number of weeks.

We designed an algorithmic core of an integrated decision support system tailored to assist farmers in managing their planting and harvesting activities. This algorithmic core is a composition of the three scheduling algorithms and one weather forecast processing algorithm. With their synergy, the designed system would generate schedules that facilitate continuous harvest, with crops harvested at their optimal maturity. Furthermore, the system enables dynamic adjustment of harvesting schedules based on actual weather conditions on the fields, thereby enhancing schedule reliability and introducing stability into further operational planning. Our initial results indicate that with the generated harvest schedules it is possible to reduce the total amount of yield exceeding stated storage and processing capacities.

3 - Organic Farming Transitions: a Dynamic Bioeconomic Model

Michael Meneses, Cornell University, Ithaca, NY, United States, Clare L. Casteel, Miguel I. Gómez, David R. Just, Ravi Kanbur, David R. Lee, C.-Y. Cynthia Lin Lawell

We develop a dynamic bioeconomic model of a farmer’s decisions regarding the use of synthetic compounds (e.g., synthetic fertilizers and pesticides) and the transition from conventional to organic management. Our crop production model accounts for newly documented interrelationships among synthetic compound use, soil health, and crop yields. In particular, new insights from soil science show that the use of synthetic compounds can be harmful to beneficial soil microbes that improve agricultural yields by enhancing crop nutrient use, stress tolerance, and pest resistance. We characterize and solve for a “fully informed” farmer’s optimal synthetic compound use strategy, and for whether and how a farmer should transition from conventional to organic farming. These solutions are compared to those from a “misinformed” model in which the farmer is not aware of the interactions between synthetic compound use, soil health, and crop yields, allowing us to assess how gaining knowledge of these interactions might be expected to change farmers’ synthetic compound use strategies and, ultimately, their decisions around adopting organic management. We identify and discuss agricultural and economic conditions under which farmers can be expected to voluntarily reduce their reliance on synthetic compounds, and possibly even adopt organic management, upon learning of the benefits associated with stewardship of their soil’s microbiome. We apply our model to farmer-level pesticide-use panel data to estimate parameters governing farmers' current understanding of the interrelations between soil microbes, pesticides, and crop yields, and to examine possible effects of extension programs targeting farmers' understanding of soil microbes.

4 - Exploring Mega-Trends and Enhancing Food Supply Chain Viability: A Global Perspective

Shanshan Yang, A*STAR, Singapore, Singapore, Singapore, Xingyu Li, Wei Ye, Cheng Yin Lee, Wenda Chen, Ning Liu

The performance of food supply chains is influenced by various factors, including seasonal demand variations, perishability, geopolitics conflicts, economic conditions, external disruptions, and regulatory changes, which collectively shape market trends and supply chain dynamics, and pose challenges to managing supply chain viability. For instance, seasonal demand shifts and economic fluctuations can strain inventory levels and logistics capacities, while geopolitical conflicts can disrupt material and ingredient supplies and increase risk exposure. Understanding mega-trends in food supply chains is crucial for gauging demand and mitigating material shortages and production disruptions.

This study aims to identify global mega-trends and their complex interactions across diverse industrial sectors and geographical regions. Leveraging the Purdue GTAP database, a base model of the global food supply chain network is constructed, encompassing demand profiles, production capabilities, and supplier-buyer relationships. Subsequently, leveraging the Mintel database, regional and sectoral food market trends are systematically analyzed, elucidating their interconnections within the constructed supply chain network. By harnessing the insights gleaned from identified mega-trends, food supply chain vulnerabilities across temporal, geographical, and industrial contexts will be evaluated under various disruption scenarios. The ultimate objective is to equip the decision-maker with the requisite knowledge to proactively navigate mega-trends and disruptions. In an industrial use case, we will demonstrate how to incorporate identified mega-trends with internal operational data, tailoring specific plans to assist companies in enhancing their operational resilience, identifying alternative products and suppliers, building redundancy, and improving demand forecasting, all of which contribute to sustained viability amid the dynamics of global supply chains.

5 - Digital Technology-Enabled Quality Bunching and Financing in Agricultural Supply Chains

Penghui Guo, School of Management,Xi'an Jiaotong University, Xi'an 710049, China, People's Republic of, Gengzhong Feng, Kai Wang

Platforms are increasingly involved in agricultural supply chains (ASCs) by providing digital technology (DT) to guide farming cooperatives' operations from the initial planting stage and by offering financing services to support agricultural production. Due to the intrinsic uncertainty in agricultural product growth, the distinctive feature of DT that we capture in ASCs, which differs from existing studies on its role in improving product quality homogeneously, is clustering agricultural product quality ranges by real-timely monitoring cultivations, i.e., the quality bunching effect. To explore how the interactions between DT implementation and platform financing affect ASC participants' decisions, we consider an ASC containing a farming cooperative and a platform that procures agricultural products and sells them to online customers. Intuitively, as the quality bunching effect increases, the high-quality aggregation of agricultural products can encourage the platform's DT provision due to increased consumer consumption. Surprisingly, we find that the platform is willing to provide DT only when the quality bunching effect is relatively low because the increased consumption raises the farming cooperative's production financing costs, in turn increasing wholesale prices and negatively affecting the platform's profits. Interestingly, we find that the farming cooperative chooses platform financing even though it faces a bank interest rate that is lower than the platform interest rate. Finally, we identify the equilibrium results for DT implementation and financing strategy and provide several extensions to consolidate our conclusions. Our findings shed light on how platforms work with farming cooperatives in agriculture digitization and financing.

6 - Optimizing Agricultural Supply Chains Using Resource-Constrained Shortest Path

Steven Formaneck, Minnesota State University Mankato, Mankato, MN, United States

The goal for most businesses is to satisfy its customers while minimizing total cost or maximizing profit. Other goals such as sustainability and risk minimization, are also common. Constraints also exist within and outside the organization that increase its costs or limit its profitability such as financial budgets, scarcity of raw materials, labor, time, and production capacity. The research provides a framework for empirically deciding what organizational decisions should be made within an agricultural organization's supply chain and in what sequence to optimize profitability given its resource constraints. The research accomplishes this through a comparative study of the leading algorithms for the resource-constrained shortest path and organizational and resource factors involved, to develop the framework. An example of the framework is provided through means of a case study.

7 - Solving a stochastic logistics network design for fresh food using a sample-average approximation

Harol Gamez Alban, Massachusetts Institute of Technology, Cambridge, MA, United States, Chris Mejia

One of the most challenging tasks in today's food industry is controlling product quality throughout the food supply chain. One of the main reasons that accelerates the spoilage of fresh food is disruptions in the supply chain network, such as poor food handling in facilities, transportation problems, handling, packaging and storage facility issues and others. Models are needed to determine the optimal supply chain configuration to maintain food freshness and, thus, reduce food waste/losses. This paper integrates food quality into the supply chain network by tracking the remaining shelf lifetime. We formulate a mixed-integer linear programming (MILP) model under uncertainty that accounts for the disruptions along the logistics network. To model and solve this problem, we use a two-stage stochastic programming with recourse and sample average approximation (SAA) method. To accelerate the convergence of the algorithm, we set up a set of valid inequalities along with a splitting procedure that speeds up the performance of the model to obtain feasible and near-optimal solutions. The resulting model is applied in an illustrative case study and can be used for planning and operating food distribution systems, considering freshness and cost. The results show that the proposed approach can achieve high-quality solutions with guaranteed performance.

TB14

Summit - 334

Machine Learning in Finance II

Invited Session

Finance

Chair: Nan Chen, Chinese University of Hong Kong, Shatin N T, Hong Kong

1 - Aico: Model-Agnostic Feature Significance Tests for Supervised Learning

Kay Giesecke, Stanford University, Stanford, CA, United States

Supervised learning algorithms are increasingly being used to guide economic and social decisions. The opaqueness of these algorithms is a key challenge, especially in highly regulated domains including financial services, where transparency is crucial. This paper provides statistical tools to improve model transparency. We develop a family of hypothesis tests for testing the significance of the input features (variables) of a supervised learning algorithm. The key idea is to evaluate a feature's incremental effect on model predictive or decision performance, relative to a baseline that does not permit variability of any feature. A null-feature's incremental effect has a distribution across a test sample whose center is non-positive. This formulation leads to simple, assumption-light uniformly most powerful sign and other tests that deliver exact, non-asymptotic p-values. These tests are model-agnostic and apply to any regression or classification model, including ensembles. They do not require retraining or other computationally costly procedures. A byproduct of the tests are interpretable measures of feature importance including exact confidence intervals that are robust to correlation among features. A hierarchical formulation can discover higher-order structure tied to subsets of features including feature interactions. A simulation study is used to demonstrate the performance of our tests and alternative approaches. An empirical study tests feature significance for a deep learning ensemble model of US home prices.

2 - Multi-task Dynamic Pricing in Credit Market with Contextual Information

Renyuan Xu, University of Southern California, Los Angeles, CA, United States, Adel Javanmard, Jingwei Ji

We study the dynamic pricing problem faced by a broker that buys and sells a large number of financial securities in the credit market, such as corporate bonds, government bonds, loans, and other credit-related securities. One challenge in pricing these securities is that most are infrequently traded, resulting in insuf-
ficient data for individual pricing. However, these products often have more structural features to utilize. To address this, we propose a multi-task dynamic pricing framework with learning, which exploits the shared structures across different securities to enhance pricing accuracy. Our approach consists of two stages: an initial regression using aggregated data across securities, followed by a refined estimation for each individual security. Theoretically we characterize the regret of our policy, which is better than both a policy that treats each security individually and a policy that treats all securities as the same. Empirical results on financial datasets show that our framework achieves better performance compared to benchmark methods.

This is based on joint work with Adel Javanmard and Jingwei Ji (USC).

3 - On Consistency of Signature Using Lasso

Chaoyi Zhao, Peking University, Beijing, China, People's Republic of, Xin Guo, Binnan Wang, Ruixun Zhang

Signatures are iterated path integrals of continuous and discrete-time processes, and their universal nonlinearity linearizes the problem of feature selection in time series data analysis. This paper studies the consistency of signature using Lasso regression, both theoretically and numerically. We establish conditions under which the Lasso regression is consistent both asymptotically and in finite sample. Furthermore, we show that the Lasso regression is more consistent with the Itô signature for time series and processes that are closer to the Brownian motion and with weaker inter-dimensional correlations, while it is more consistent with the Stratonovich signature for mean-reverting time series and processes. We demonstrate that signature can be applied to learn nonlinear functions and option prices with high accuracy, and the performance depends on properties of the underlying process and the choice of the signature.

TB15

Summit - 335

Recent advances in Choice Modeling

Invited Session

Revenue Management and Pricing

Chair: Ashwin Venkataraman, University of Texas at Dallas, Richardson, TX, United States

1 - Assortment Optimization with Visibility Constraints

Marouane Ibn Brahim, Cornell University, New York, NY, United States, Théo Barré, Omar El Housni, Andrea Lodi, Danny Segev

Motivated by applications in e-retail and online advertising, we study the problem of assortment optimization with visibility constraints, termed AOV. Given a universe of substitutable products and a stream of T customers, the objective is to determine the optimal assortment of products to offer to each customer in order to maximize the total expected revenue, while ensuring an exogenously-given minimum display requirement, referred to as visibility constraints. We assume that customer choices follow a Multinomial Logit model. We provide a characterization of the structure of the optimal assortments and present an efficient polynomial time algorithm for solving AOV. Our algorithm leverages the supermodularity of a novel function called "expanded revenue" of an assortment. Additionally, we demonstrate that AOV can be formulated as a compact linear program. Next, we consider AOV with cardinality constraints, which we prove to be strongly NP-hard and not admitting an FPTAS, even when all products have identical prices. Subsequently, we devise a PTAS for AOV under cardinality constraints with identical prices. Our algorithm linearizes the objective function through a carefully crafted guessing procedure, then solves the linearized program, and randomly rounds its solution to derive a near optimal solution. We also examine the revenue loss due to visibility constraints, and we propose a strategy to distribute the loss among the products subject to visibility constraints. Each vendor is charged an amount proportional to their product's contribution to the revenue loss. Finally, we present the results of our numerical experiments providing illustration of the obtained outcomes.

2 - Assortment and Price Optimization Under an Endogenous Context-Dependent Multinomial Logit Model

Yicheng Bai, Amazon, New York, NY, United States, Omar El Housni, Paat Rusmevichientong, Huseyin Topaloglu

Motivated by empirical evidence that the utility of each product depends on the assortment of products offered along with it, we propose an endogenous context-dependent multinomial logit model (Context-MNL) under which the utility of each product depends on both the product's intrinsic value and the deviation of the intrinsic value from the expected maximum utility among all the products in the offered assortment. Under the Context-MNL model, an assortment provides a context in which customers evaluate the utility of each product. Our model generalizes the standard multinomial logit model and allows the utility of each product to depend on the offered assortment. The model is parsimonious, requires only one parameter more than the standard multinomial logit model, captures the assortment-dependent effect endogenously, and does not require the decision-maker to determine in advance the relevant attributes of the assortment that might affect the product utility. The Context-MNL model also admits tractable maximum likelihood estimation and is operationally tractable, with efficient solution methods for solving assortment and price optimization problems. Our numerical study, which is based on data from Expedia, shows that compared to the standard multinomial logit model, the Context-MNL model substantially improves out-of-sample goodness of fit and prediction accuracy.

3 - A Simple Way Towards Fair Assortment Planning: Algorithms and Welfare Implications

Wentao Lu, Carey Business School, Johns Hopkins University, Baltimore, MD, United States, Ozge Sahin, Ruxian Wang

Traditional assortment optimization creates unfairness among sellers as only a single assortment with the highest expected revenue is chosen resulting in some sellers being excluded from recommendations or assortments, with minimal market exposure and revenue. To address this issue, we incorporate fairness constraints that ensure fair market exposures for all sellers. These constraints ensure each seller to have a minimum market exposure, which may depend on the seller's reputation, product quality, and price, among other features. We show that the optimal solution with fairness constraints is to randomize over at most n nested assortments, where n is the number of sellers (or products), and the optimal solution can be found in polynomial time. For cases in which there are other business constraints including imposing cardinality constraints on the assortments and limiting the number of different assortments, we characterize the structure of the optimal solutions and propose efficient heuristics. We further investigate the impact of fairness constraints on consumer welfare, and show that it always increases when such constraints are imposed. Our analysis reveals that when fairness constraints induce new sellers with high quality products to enter the platform, all involved parties are better-off resulting in a win-win-win situation. Even when there is no new seller entry, we identify cases in which the total welfare improves and therefore propose a revenue redistributing mechanism to achieve a win-win-win solution.

4 - Generalized Stochastic Preference Choice Model

Ashwin Venkataraman, University of Texas at Dallas, Richardson, TX, United States

We propose a new discrete choice model, called the generalized stochastic preference (GSP) model, that incorporates non-rationality into the stochastic preference (SP) choice model, also known as the rank-based choice model. Our model can explain several choice phenomena that cannot be represented by any SP model such as the compromise and attraction effects, but still subsumes the SP model class. The GSP model is defined as a distribution over consumer types, where each type extends the choice behavior of rational types in the SP model. We build on existing methods for estimating the SP model and propose an iterative estimation algorithm for the GSP model that finds new types by solving a integer linear program in each iteration. We further show that our proposed notion of non-rationality can be incorporated into other choice models, like the random utility maximization (RUM) model class as well as any of its subclasses. As a concrete example, we introduce the non-rational extension of the classical MNL model, which we term the generalized MNL (GMNL) model and present an efficient expectation-maximization (EM) algorithm for estimating the GMNL model. Numerical evaluation on real choice data shows that the GMNL and GSP models can outperform their rational counterparts in out-of-sample prediction accuracy.

TB16

Summit - 336

Design and Operation of Online Platforms

Invited Session

Revenue Management and Pricing

Chair: Hongyao Ma, Columbia University, New York, NY, 10027, United States

Co-Chair: Francisco Castro, UCLA Anderson School of Management, Los Angeles, CA, United States

Co-Chair: Chiwei Yan, University of California Berkeley, San Francisco, CA, United States

1 - Computing Equilibrium for a Class of Dynamic Queueing Games

Chiwei Yan, University of California Berkeley, San Francisco, CA, United States, Francisco Castro, Hongyao Ma, Denise Cerna

We will discuss a general dynamic equilibrium solver for a class of dynamic queueing games involving strategically accepting jobs. It has wide applications in many market design applications, including kidney allocation, ridesharing matching, etc. We discuss a simple procedure for equilibrium computation and discuss its convergence properties. We showcase promising empirical results based on ridesharing airport dispatching data that uncover interesting new insights that have not appeared in our previous fluid analysis (Castro, Ma, Nazerzadeh, and Yan 2021).

2 - Who Ate the Lunch? a Variational Em Algorithm for Identifying Refund Hunters

Chenkai Yu, Columbia University, New York, NY, United States, Arpit Agarwal, Hongyao Ma

For online platforms such as UberEats and DoorDash, a central challenge in customer service is the lack of ground truth — when a customer reports that an order was never delivered, it is difficult for support agents to determine if the driver kept the food, a passer-by took the delivery, or the customer is falsely claiming a missing order. This uncertainty often results in platforms shouldering the costs for refunds, without holding either side of the market accountable. In this work, we propose a variational Expectation-Maximization (EM) algorithm for evaluating the trustworthiness of each customer and driver, considering both their frequency of platform use and the reliability of the agents (on the other side of the market) each participant interacted with. We show via simulations that our algorithm significantly improves accuracy in identifying strategic participants, compared to several baseline algorithms, including a naive method that scores each participant based only on their past defect rates.

3 - Information Sharing Effects of High Stock Delivery Windows on Online Platforms

Do Yoon Kim, Boston College, Chestnut Hill, MA, United States, Benjamin Knight, Dmitry Mitrofanov

In this paper, we analyze a novel information-sharing policy to reduce stockouts on online grocery retail platforms and thereby improve platform performance. This policy involves leveraging the digital interface of online platforms to disseminate information about delivery windows when retailer stock levels are at their peak. Ex-ante, it is unclear if such a policy would positively influence customers' purchasing patterns, particularly in terms of its impact on platforms' fundamentals. We conducted a large-scale field experiment on over 1M users on Instacart, an online grocery delivery platform. In this experiment, customers in the treatment group were prompted "Higher stock at this time" under delivery windows corresponding to peak supermarket stock levels. Interestingly, we find that this cost-free information-sharing policy significantly increased customers' conversion rates, as well as the propensity of customers to pick high-stock delivery windows. We also observe that this policy reduces item refund and replacement rates by 2.2% and 2.7%, respectively, as well as increases the platform's revenue by 6\% and the number of orders placed by 3%. Finally, treated users increase orders from stores they have not previously ordered from and order novel items, suggesting an stock information has an exploratory effect. Our findings suggest simple information sharing policies can increase customer orders, enhance their willingness to explore new products and retail partners, and potentially lead to a more significant positive long-term impact on platform fundamentals.

4 - Improving Worker Learning in the Gig Economy

Shunan Jiang, University of California, Berkeley, Berkeley, CA, United States, Park Sinchaisri

The gig economy is experiencing exponential growth, with an increasing number of workers engaging in on-demand platforms across various sectors, from grocery delivery to healthcare. These gig workers, often independent contractors, face unique challenges in their work environment as they lack structured training programs and supervision. Therefore, understanding how these workers learn and improve their performance becomes crucial. In collaboration with a grocery delivery platform, our work aims to explore the decision-making and learning processes employed by on-demand shoppers and inform the development of effective training and scheduling programs tailored to the unique needs of gig economy workers.

TB17

Summit - 337

Data Science and Artificial Intelligence for Business

Invited Session

Revenue Management and Pricing

Chair: Yifan Feng, National University of Singapore, Singapore, Singapore

1 - Optimizing Scalable Targeted Marketing Policies with Constraints

Yuting Zhu, National University of Singapore, Singapore, Singapore, Haihao Lu, Duncan Simester

Targeted marketing policies target different customers with different marketing actions. While most research has focused on training targeting policies without managerial constraints, in practice, many firms face managerial constraints when implementing these policies. For example, firms may face volume constraints on the maximum or minimum number of actions they can take, or on the minimum acceptable outcomes for different customer segments. They may also face similarity (fairness) constraints that require similar actions with different groups of customers. Traditional optimization methods face challenges when solving problems with either many customers or many constraints. We show how recent advances in linear programming can be adapted to the targeting of marketing actions. We provide a theoretical guarantee comparing how the proposed algorithm scales compared to state-of-the-art benchmarks (primal simplex, dual simplex and barrier methods). We also extend existing guarantees on optimality and computation speed, by adapting them to accommodate the characteristics of targeting problems. We implement the proposed algorithm using data from a field experiment with over 2 million customers, and six different marketing actions (including a no action “Control”). We use this application to evaluate the computation speed and range of problems the algorithm can solve, comparing it to benchmark methods. The findings confirm that the algorithm makes it feasible to train large-scale targeting problems that include volume and similarity constraints.

2 - Choice Model Estimation with Stockout-Based Substitution in Censored Demand Processes

Longlin Wang, Harvard University, Cambridge, MA, United States, Ningyuan Chen

In brick-and-mortar retailing, shelves are stocked with multiple units of products. Customers arrive during the day and make purchases governed by a discrete choice model and stockout-based substitution. The store in many cases only has access to the inventory level of each product at the beginning and the end of the horizon, while individual decisions in the process are undisclosed. We study the identifiability of the problem and propose a method to estimate parameters of a multinomial logit (MNL) choice model in this highly censored situation. The proposed estimators are empirical risk minimizers solved by a stochastic first-order algorithm. Our work adds to the estimation perspective of prevalent MNL-based dynamic assortment planning problems. We showcase the efficacy of our method with both synthetic experiments and industry applications where our learning task indeed arises.

3 - Feature Misspecification in Sequential Learning Problems

Dongwook Shin, HKUST Business School, Clear Water Bay, Hong Kong, Dohyun Ahn, Assaf Zeevi

We consider a class of sequential learning problems, where a decision maker must learn the unknown statistical characteristics of a finite set of alternatives (or systems) using sequential sampling to ultimately select a subset of "good" alternatives. A salient feature of our problem is that system performance is governed by a set of features. The decision maker postulates the dependence on these features to be linear, but this model may not precisely represent the true underlying system structure. We show that this misspecification, if not managed properly, can lead to suboptimal performance due to a phenomenon identified as sample-selection endogeneity. We propose a prospective sampling principle—a new approach that eliminates the adverse effects of misspecification as the number of samples grows large. The proposed principle applies across a very general class of widely used sampling policies, enjoys strong asymptotic performance guarantees, and exhibits effective finite-sample performance in numerical experiments.

4 - A Behavioral Model for Multi-Armed Bandits: Theoretical Framework and Experimental Evidence

Jingying Ding, National University of Singapore, Singapore, Singapore, Yifan Feng, Ying Rong

The exploration-exploitation trade-off is a fundamental concept arising when the decision maker needs to make repeated choices, whose rewards are unknown *in priori*. This paper delves into how humans navigate this trade-off through the lens of the multi-armed bandit (MAB) problem. Inspired by both behavioral economics and online learning literature, we introduce a novel family of behavioral policies called Myopic Quantal Choice (MQC). It is a dynamic adaptation of quantal choice models, deriving anticipated utilities directly from past rewards. MQC offers a simple method to describe the arm selection process, and yet is rich enough to quantify the exploration-exploitation trade-off through a "shrinkage rate of exploration."
Through both non-asymptotic and asymptotic analysis, we show that MQC admits intuitive properties that match the qualitative patterns of the laboratory experiment data. Particularly, MQC always converges to the optimal arm, thus capturing the "learning" effect. In addition, we characterize MQC's regret in the asymptotic regime and demonstrate the effects of "over-" and "under-" exploration; Over-exploration with a too small shrinkage rate parameter results in gradual deterioration of the lower bound of regret, while under-exploration with a too large rate parameter leads to sudden deterioration. Analysis of laboratory experiment data reveals that the MQC model excels in predictive power compared to other behavioral models. Insights from the asymptotic regime also extend to the finite horizon experiments. Particularly, when we fit the MQC model to the data, a prevalent tendency toward over-exploration becomes evident.

TB18

Summit - 338

Human-Algorithm Interactions in Behavioral Analytics

Invited Session

Revenue Management and Pricing

Chair: Yueyang Zhong, London Business School, London, N/A

1 - Learning when to Advise Human Decision Makers

Gali Noti, Cornell University, Ithaca, NY, United States, Yiling Chen

<blockquote mathjax="">Artificial intelligence (AI) systems are increasingly used for providing advice to facilitate human decision making in a wide range of domains, such as healthcare, criminal justice, and finance. Motivated by limitations of the current practice where algorithmic advice is provided to human users as a constant element in the decision-making pipeline, in this paper we raise the question of when should algorithms provide advice? We propose a novel design of AI systems in which the algorithm interacts with the human user in a two-sided manner and aims to provide advice only when it is likely to be beneficial for the user in making their decision. The results of a large-scale experiment show that our advising approach manages to provide advice at times of need and to significantly improve human decision making compared to fixed, non-interactive, advising approaches. This approach has additional advantages in facilitating human learning, preserving complementary strengths of human decision makers, and leading to more positive responsiveness to the advice.

2 - Information Aggregation In Marketplaces

Dragos Florin Ciocan, INSEAD, Fontainebleau, France

Modern marketplaces function not only as facilitators of transactions but also as information aggregators. This leads to information asymmetry where the marketplace itself may have superior knowledge of demand than the sellers transacting through it. We investigate optimal sale mechanisms in this situation.

3 - A Space-Efficient Algorithm for Counting Distinct Elements in the Turnstile Model

Tingting Ou, Columbia University, New York, NY, United States, Rachel Cummings, Alessandro Epasto, Jieming Mao, Tamalika Mukherjee, Peilin Zhong

The continual observation streaming setting captures data analysis scenarios where both data collection is ongoing, and real-time analysis of the collected data is needed at all time points. The \emph{turnstile} streaming model is a subclass of the continual observation model and allows for both the insertion and deletion of data items over time. Typically, the length of the stream T and the size of the universe $\cU$ from which data items come from are considered to be extremely large, thus practical solutions should use space at most sublinear in T and $\mathcal{U}$. Because many of the relevant applications for data analysis in the continual observation setting involve the analysis of sensitive user data, this motivates the need to develop space-efficient algorithms that achieve formal privacy protections such as differential privacy.

We give the first sublinear space differentially private algorithm for the fundamental problem of counting distinct elements in the turnstile streaming model that achieves $\Tilde{O}(T^{1/3})$ space and additive error and a $(1+\eta)$-relative approximation for all $\eta \in (0,1)$. Our algorithm significantly improves upon the space requirements of the state of the art for this problem in the turnstile model~\cite{JainKRSS23} which has a linear dependency in both T and $\vert \cU \vert$, while still achieving an additive error that is close to the known $T^{1/4}$ lower bound.

4 - Human Expertise in Algorithmic Prediction

Rohan Alur, Massachusetts Institute of Technology, Cambridge, MA, United States

We introduce a novel framework for incorporating human expertise into algorithmic predictions. Our approach focuses on the use of human judgment to distinguish inputs which `look the same' to any feasible predictive algorithm. We argue that this framing clarifies the problem of human/AI collaboration in prediction tasks, as experts often have access to information---particularly subjective information---which is not encoded in the algorithm's training data. We use this insight to develop a set of principled algorithms for selectively incorporating human feedback only when it improves the performance of any feasible predictor. We find empirically that although algorithms often outperform their human counterparts *on average*, human judgment can significantly improve algorithmic predictions on *specific* instances (which can be identified ex-ante). In an X-ray classification task, we find that this subset constitutes nearly 30% of the patient population. Our approach provides a natural way of uncovering this heterogeneity and thus enabling effective human-AI collaboration.

TB19

Summit - 339

Economic Decision-Making: Games and Simulations

Contributed Session

Chair: Guilherme Tortorella, University of Melbourne, Docklands, Australia

1 - An Experimental Investigation of Bounded Rationality of a Newsvendor Under Supply Shortages

Anannya Gogoi, Rajagiri Business School, Kochi, India, Ashutosh Sarkar

This study aims to examine how retailers react to supply disruptions, which contribute to additional complexity during inventory decision-making. We explore the shortage gaming exhibited by retailers in a single supplier and multiple retailers supply chain. We conduct a between subjects behavioral experimental game to capture the retailer's sub-optimal decision-making under uncertain supply shortages with four experimental sessions. The experimental sessions were in a 2 (capacity uncertainty) X 2 (magnitude of disruption) factorial design. We adopt the Quantal Response Equilibrium (QRE) framework to explain the behavioral irregularities. Our results reveal the presence of bounded rationality in retailers during decision-making under supply disruption across all the treatment conditions. The results show that supply disruption significantly affects retailers' ordering behavior. The probability of supply disruption affects the ordering decisions significantly. Finally, we discuss how these results may help researchers and practitioners better understand retailers' ordering behavior and provide implications considering disruptions at the supplier's end.

Keywords: *Behavioral operations, Experiments, Newsvendor model, Shortage gaming, Supply disruption.*

2 - A Gameful Learning Approach in Engineering Economics

Esra Agca Aktunc, Rensselaer Polytechnic Institute, Troy, NY, United States

An undergraduate Engineering Economics course is redesigned with gameful and active learning interventions to increase student engagement and achieve higher-level learning outcomes including analysis, synthesis, and evaluation. The main additional activities are CasePlays which are case studies that require collaborative problem solving and in-person role-playing from teams of students. CasePlays include tasks of formulating cash flows, evaluating alternatives based on performance measures such as rate of return, carrying out cost-benefit analysis, replacement analysis, sensitivity analysis, or portfolio analysis, and making recommendations. CasePlays involve random disruptions and require discussions among different stakeholders such as project managers, investors, government regulators, or environmental activists. Students are expected to consider ethical, environmental, sustainability, and societal factors in decision making as a team. Following the in-person activity, students are expected to reflect on their learning experience through an individual written response. The assessment strategies are modified by using rubrics and changes in grading policies. The student performance results, feedback, and observations based on a pilot study of the redesigned course will be presented.

3 - A Card Game for Teaching TCE-Based Make-or-Buy Decisions

Beverly Osborn, Indiana University, Indianapolis, IN, United States

I developed this card game to use in the make-or-buy section of my in-person class on global sourcing. It gives students an opportunity to discover and discuss concepts of asset specificity, opportunism, bargaining power, signaling, contract management and incomplete contracts. I will demonstrate the game, share my teaching materials, and request feedback. If time and attendance allow, we will play a round of the game during the session.

TCE Card Game Overview: Each student is assigned the role of either a buyer or a supplier. Students play the game in buyer-supplier pairs and aim to earn as many points as possible. Buyers compete only against other buyers, and suppliers compete only against other suppliers. Each player's decisions affect both their own points and the points of their supply chain partner. Opportunistic behavior tends to arise in line with theoretical predictions.

4 - Artificial Intelligence and Quality Management

Guilherme Tortorella, University of Melbourne, Docklands, Australia, Daniel Nascimento

This study empirically examines the effects of artificial intelligence (AI) integration into quality management (QM) processes of service organizations. A multi-case study in service organizations of different types was conducted. We utilized multiple sources of evidence to analyze and triangulate findings, yielding the development of propositions in light of the Dynamic Capabilities theory. Our findings indicated that AI adoption positively supports QM, especially in terms of the organization's ability to quickly learn and create new quality control and assurance solutions/procedures from this learning (sensing), and to integrate them into organizational processes and routines (seizing). The identification of the impact of AI adoption on QM helps service organizations anticipate potential issues and address adequate countermeasures to reinforce their QM processes towards superior performance results.

TB20

Summit - 340

Innovative Decision-Making in Dynamic Environments: Trends and Applications

Invited Session

Decision Analysis Society

Chair: Zheng Zeng, Rutgers University, East Brunswick, NJ, United States

1 - Drivers of Reshoring and Offshoring Decisions in the Life Sciences Industry

Aman Goswami, Rutgers Business School, Newark, NJ, United States, Alok Baveja, Esther Ho, Lian Qi, Jing Wang

Reshoring and offshoring are considered strategic approaches to sourcing and managing production operations, offering significant benefits to a firm's operations, costs, and risks. In this research, we examine the joint drivers of reshoring and offshoring decisions for life sciences firms based in the USA. This research develops a conceptual framework based on extant literature and compiles empirical panel data curated from secondary sources. We propose and validate hypotheses with models that link economic factors (e.g., GDP), firm-specific factors (e.g., tax rates), industry-specific factors (e.g., industry type), and risk factors (e.g., exchange rate fluctuations) to reshoring and offshoring decision-making. Further, the amount of capital investment and projected jobs resulting from such decisions are also analyzed. Our findings uncover the importance of the overall economic development, market size, purchasing power, and the cost of doing business in a country as important drivers of reshoring and offshoring decisions. We also find that the firm size moderates the reshoring and offshoring decision-making, with larger firms presenting stronger reshoring capabilities.

2 - O2O On-Demand Delivery: Order Assignment with Behavior Heterogeneity of Crowd-Sourced Drivers

Xinyi Huang, University of Iowa, Iowa City, IA, United States, Tonghua Lin, Weiwei Chen, Hongyan Dai

Online-to-Offline (O2O) on-demand services, such as food and grocery delivery platforms, enable customers to place orders online and receive their products offline shortly thereafter. O2O customers typically have high expectations for delivery times, leading many platforms to rely heavily on crowdsourced drivers for order fulfillment. However, these platforms often fail to recognize the heterogeneity of drivers and neglect to consider their varying routing preferences. Therefore, the algorithm might mismatch the drivers' capacities with the complexities of orders and ignore the drivers' local knowledge. We have also observed disparities between drivers' assigned routes and their actual routing choices in our data. We aim to integrate routing predictions into our optimization. By incorporating the learned heterogeneity of crowdsourced drivers, we intend to enhance the order assignment and sequencing algorithm. Our ultimate goal is to establish a "human-in-the-loop" framework that aligns the algorithm with real-world scenarios, enabling a more effective balance between algorithms and human input.

3 - SpatialRank: Urban Event Ranking with NDCG Optimization on Spatiotemporal Data

Bang An, University of Iowa, Iowa City, IA, United States, Xun Zhou

The problem of urban event ranking aims at predicting the top-k most risky locations of future events such as traffic accidents and crimes. This problem is of fundamental importance to public safety and urban administration especially when limited resources are available. The problem is, however, challenging due to complex and dynamic spatio-temporal correlations between locations, uneven distribution of urban events in space, and the difficulty to correctly rank nearby locations with similar features. Prior works on event forecasting mostly aim at accurately predicting the actual risk score or counts of events for all the locations. Rankings obtained as such usually have low quality due to prediction errors. Learning-to-rank methods directly optimize measures such as Normalized Discounted Cumulative Gain (NDCG), but cannot handle the spatiotemporal autocorrelation existing among locations. In this paper, we bridge the gap by proposing a novel spatial event ranking approach named SpatialRank. SpatialRank features adaptive graph convolution layers that dynamically learn the spatiotemporal dependencies across locations from data. In addition, the model optimizes through surrogates a hybrid NDCG loss with a spatial component to better rank neighboring spatial locations. We design an importance-sampling with a spatial filtering algorithm to effectively evaluate the loss during training. Comprehensive experiments on three real-world datasets demonstrate that SpatialRank can effectively identify the top riskiest locations of crimes and traffic accidents and outperform state-of-art methods in terms of NDCG by up to 12.7%.

4 - Patient-Centric Appointment Scheduling Considering Patient Heterogeneity

Chenhao Zhou, Rutgers Business School, Newark, NJ, United States, Weiwei Chen, Xin Ding, Lei Lei

Deterministic models traditionally used for resource allocation are inadequate for the practical realities of scheduling chemotherapy outpatient visits, particularly at the Cancer Institute of New Jersey (CINJ), where patient preferences and appointment timings can vary unpredictably. This limitation necessitates a more sophisticated approach that accommodates inherent variability and uncertainty in patient scheduling. This research proposes a two-stage scheduling approach designed to address these limitations, aiming to enhance both patient satisfaction and operational efficiency by incorporating patient preferences and real-time resource availability.

In Stage I, a recommender system is introduced to assign patients to their preferred appointment windows during the call. This stage balances operational efficiency with patient uncertainties through three key components: Patient Late/Early Identification, which uses historical data to predict the likelihood of patients arriving as scheduled; Waiting Time Prediction, which estimates wait times from patient check-in to treatment initiation to recommend optimal time windows; and a Model-based Recommender System utilizing deep reinforcement learning to dynamically adjust recommendations based on ongoing patient bookings and resource availability.

Stage II involves refining the allocation of specific timeslots within the preferred windows as the appointment date approaches. This phase employs a stochastic model to adapt to last-minute changes and uncertainties, representing a significant advancement over the deterministic model previously used. By addressing the dynamic and uncertain nature of appointment scheduling at CINJ, this research aims to enhance patient satisfaction and operational efficiency, ultimately improving the alignment between patient needs and resource availability.

5 - An Empirical Study on the Impact of Social Media User-Generated Policy Narratives on Citizen Engagement Behaviors

Min Zhang, Communication University of China, Beijing, China, People's Republic of, Fei Wang

The importance of citizen engagement in policy communication has grown with the advent of social media, which offers novel avenues for public interaction that transcend traditional media limitations. In this context, the study investigates the transformative potential of personalized policy narratives on social media platforms to engage the public more deeply in policy discourse. This study examines the impact of personalized policy narratives on public engagement through social media, challenging the traditional focus on standardized narrative content within the narrative policy framework. Using narrative transportation theory, the research analyzes 10,577 short videos from the Douyin platform related to new energy vehicle policies. Employing a combination of deep learning, dictionary-based approaches, and manual coding, the study identifies three key characteristics of policy narratives: cognitive embedding, affective embedding, and reference embedding. It explores how these characteristics influence public participation behaviors—specifically likes and shares—and examines the moderating effects of policy type and narrator influence. The findings reveal that each narrative characteristic differentially impacts engagement, with contextual factors significantly shaping these effects. This paper contributes to narrative policy studies by introducing advanced methodologies for analyzing large-scale social media data and provides new insights into the role of personalized narratives in enhancing public engagement with policy issues.

TB21

Summit - 341

Decision Models for Network Security and Resilience

Invited Session

Decision Analysis Society

Chair: Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States

1 -

Strategic Coordination of Heterogeneous Resources for Attack Detection

Pichamon Anukulkarnkusol, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan

We study a two-player network inspection game in which the defender coordinates sensors with potentially heterogenous detection capabilities to detect attacks placed by a strategic attacker. We assume that there are multiple types of sensors, each determined by the detected area coverage when placed at each detector location. The objective of the defender (resp. attacker) is to minimize (resp. maximize) the expected number of undetected attacks. We analytically characterize Nash equilibria of this zero-sum game in scenarios where there is no overlapping detected area coverages upon placing sensors at any detector location, and both players possibly perform a randomized strategy. We then introduce a new solution approach derived from our analysis and validate its efficacy through computational experiments.

2 - Strategic Allocation of Heterogeneous Resources for Network Security

Bobak McCann, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan

We consider a network inspection game, in which a defender positions heterogeneous detectors according to a probability distribution in order to detect multiple attacks caused by a strategic attacker. We assume the defender has access to multiple types of detectors that can potentially differ in their accuracy and cost, and that the attacker has access to multiple types of attack units that can potentially differ in their effectiveness. The objective of the defender (resp. attacker) is to minimize (resp. maximize) the expected number of undetected attacks. We provide a polynomial algorithm to compute Nash Equilibria for this game under the assumption that each component in the network can be monitored from a unique detector location.

3 - Generalization bounds for sequential time-series and supply chain decision models

William Zhang, MIT Operations Research Center, Cambridge, MA, United States, Saurabh Amin, Georgia Perakis, Aron Brenner

Supply chain analytics often involves forecasting quantities (such as demand or supply of critical parts) for the purpose of deciding order quantities or for pricing over a time horizon. Here we consider a predict-then-optimize (or sequential learning and optimization) setting in which nested time-series models are learned based on multi-source contextual data, including macroeconomic indicators correlated with supply chain performance indicators. We develop new generalization bounds that account for two key characteristics: (1) Differences in the data availability of various contextual features and structural relationships; (2) Impact of forecasting accuracy on downstream decisions. We develop new generalization bounds by leveraging statistical properties of strong mixing processes and stability of learning algorithms for this setting. Our results combine and extend prior work in both generalization bounds for non-iid data and predict-then-optimize problems. We illustrate using classical examples such as the newsvendor problem and two-stage least squares. We also highlight implications for model selection for demand forecasting and pricing decisions for auto industry.

4 - Military Logistics Planning for Expeditionary Warfare

J. Haden Boone, Georgia Institute of Technology, Atlanta, GA, United States, Matthew Ford, Samuel Tan, Mathieu Dahan, Peter Frazier, Yongjia Song, Huseyin Topaloglu

Problem definition: In times of geopolitical instability, the United States Marine Corps (USMC) may be required to quickly deploy expeditionary forces to contested areas. We develop a decision-support tool that designs military logistics plans for supporting flow and sustainment of such forces in expeditionary environments. ***Methodology:*** We propose a scheduled service network design model to optimize the routing and scheduling of various transportation vehicles (connectors) and commodities across an intermodal expeditionary logistics network. Formulated as a mixed-integer program (MIP), the objective of this model is to minimize penalties incurred for late force closure and fulfillment of demand, while accounting for connector capacities, speeds, and ranges, as well as the availability of commodities. ***Results:*** We illustrate the application of our model through a fictitious warfare scenario set in Southern California designed to reflect USMC training scenarios. Our results highlight the effectiveness of the proposed MIP in creating complex intermodal logistics plans to ensure timely fulfillment of expeditionary demands. This framework offers the first optimization-based decision-support tool for expeditionary logistics planning, contrasting with the current qualitative analyses conducted by military logistics planners. Our findings provide the USMC with valuable granular and aggregated insights into the efficiency and vulnerability of their logistics network.

5 - Dynamic Operational Planning in Warfare: A Stochastic Game Approach to Military Campaigns

Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States, Joseph McCarthy, Chelsea White

We study a two-player discounted zero-sum stochastic game model for operational planning in military campaigns. At each stage, the players manage multiple commanders who order military actions on objectives that have an open line of control. When a battle occurs to determine control of an objective, its stochastic outcome depends on the actions and the enabling support provided by the control of other objectives. Each player aims to maximize the cumulative number of objectives they control, weighted by their criticality. To solve this large-scale stochastic game, we derive properties of its Markov perfect equilibria, significantly reduce state and action spaces, and accelerate Shapley's value iteration algorithm by eliminating dominated actions. We demonstrate the computational value of our equilibrium results on a case study that reflects representative operational-level military campaigns with geopolitical implications. Our analysis reveals a complex interplay between the game's parameters and dynamics in equilibrium, resulting in new military insights for strategic leadership.

TB22

Summit - 342

Emerging Technology and Decision Analysis

Invited Session

Decision Analysis Society

Chair: Qian Jia, Nanjing University, Nanjing, N/A

1 - Leveraging User Intent Extraction to Enhance Search Recommendation Efficacy in Online Shopping Platform

Li Lin, Nanjing University Business School, Nanjing, China, People's Republic of, Qianzhou Du, Alan Wang, Wenqi Shen

In the digital era, an increasing number of consumers are turning to online shopping platform for their daily needs, highlighting the imperative to comprehend consumer behavior. To assist in this endeavor, online shopping platform leverage query-based search systems in their apps, enabling customers to effortlessly locate information and products that meet their preferences. However, previous studies have rarely explored the role of user intent extraction in enhancing the effectiveness of these queries. We propose a novel deep learning model that initially classifies user queries based on their intent, distinguishing between informational and navigational purposes. Following this intent classification, we construct an improved two-tower model tailored to these distinct user intents. We validate our model using search history and click data from a renowned Chinese online shopping Platform RT-Mart. The analysis confirms that recognizing different user intents notably improves Hit Rate and Normalized Discounted Cumulative Gain(NDCG) . Finally, we demonstrate the scalability of our framework and identify the set of optimal features that maximizes accuracy.

2 - Empowering Content Creators to Add Link to Competing Platforms: a Blessing or Curse?

Jianchao Sheng, University of Science and Technology of China, Hefei, China, People's Republic of, Jianqing Chen

The digital content industry is experiencing rapid growth, with numerous platforms hosting an extensive type of content generated by independent creators. A primary revenue stream for these platforms is advertising, contingent upon the number of active customers viewing the content. The conventional advice is to raise barriers to protect consumers from competitive poaching. However, we see instances that appear to contradict this advice. One such practice is when Instagram allows its content creators to add link to competitive digital content platforms, including TikTok. In this paper, we develop a game-theoretic model to examine whether and under what conditions empowering creators to add link to competing platforms can be a beneficial strategy. We analyze two competing digital content platforms, a representing creator, and a mass of consumers. We find that when the advertising fee is too high or too low, neither platform chooses to allow creators to add link in equilibrium. Otherwise, at least one platform chooses to allow such behavior. Interestingly, when the advertising fee is moderately high, an asymmetric equilibrium in which only one platform allows such behavior can emerge.

3 - Optimal Resource Allocation Strategies for Venture Capitals in a Contest

Manman Wang, University of Science and Technology of China, Hefei, China, People's Republic of, Zihao Zhang

Many innovators compete with each other to meet the refined needs of consumers. Should venture capitals (VCs) favor a high-efficiency or low-efficiency start-up in a contest? Despite a widespread emphasis on investing more capital in the best start-up, a VC continues to invest in poor performers. We simulate the investment decisions of monopoly VC and competing VCs in competing start-ups through contest theory. We find that a monopoly VC will always invest more in inefficient start-ups and artificially reduce the success rate of efficient start-ups, thus prompting both start-ups to expend more effort. As for competing VCs, the VC with strong funds will take the initiative to invest more in inefficient start-ups under certain conditions. Our study provides some new insights for VCs on how to effectively allocate resources in highly competitive startups.

4 - Integrating LLMs and Psychometrics: Global Construct Validity

Sen Yan, University of Colorado, Boulder, Boulder, CO, United States, Kai Larsen, Roman Lukyanenko

The psychometric approach in IS offers a foundational framework for a broad spectrum of research endeavors, which typically rely on construct validation to confirm that a series of indicators accurately measures the intended construct. However, a longstanding issue with construct validity, unaddressed since its introduction by Cronbach and Meehl in 1955, is that it is evaluated using study-specific response data without comparison to constructs outside the study. This oversight (or, rather, incapability) has significant implications. We introduce a large language model combined with principal components analysis (PCA) and develop the Validity Lodestar application. This approach lays the groundwork for developing more accurate and reliable theoretical models, marking a significant leap forward in the IS discipline's methodological capabilities, making IS the first of more than a dozen psychometric disciplines with the capability to evaluate construct validity truly.

TB23

Summit - 343

AI-Driven Insights: From Operations to Financial Decision Making

Contributed Session

Chair: Hao Huang, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, Taoyuan, 32003, Taiwan

1 - Decision-Aware Predictive Model Selection for Individualized Worker Assignment

Eric Stratman, University of Wisconsin, Madison, WI, United States, Laura Albert, Justin Boutilier

Organizations depend on human decision-makers, especially in high-risk industries where information is scarce. Recognizing that a workers' experience influence their decisions, we propose a framework that integrates machine learning with integer optimization to assign tasks based on behavioral insights. Our method is a decision-aware approach that optimizes predictive model selection and data aggregation of the machine learning task for each worker to best support the downstream assignment task. We show this approach is crucial in environments characterized by complex predictions and limited data. We apply our proposed method to assign a workforce of insurance underwriter to requests for auto insurance, showcasing significant improvements in operational efficiency.

2 - Integrating Human Knowledge through Action Masking in Reinforcement Learning for Operations Management: A Case Study Analysis

Mirko Stappert, University of Freiburg, Freiburg im Breisgau, Germany, Bernhard Lutz, Niklas Goby, Dirk Neumann

Reinforcement learning (RL) provides a powerful method to address problems from operations management. However, its real-world application often fails due to a lack of user acceptance and trust. In particular, managers should be provided with possibilities of altering the learned policy by including human expert knowledge. In this study, we analyze the benefits and limitations of including human knowledge via action masking. While action masking has been used to exclude invalid actions, its potential to incorporate human knowledge and enforce specific actions has not been explored yet. We present the results of three numerical case studies focusing on the problems of peak load management, paint shop scheduling, and inventory management. Our findings suggest that using action masking to include human knowledge in the form of reasonable heuristics can lead to considerable cost reductions. Furthermore, we find that including human knowledge can be strictly necessary to learn effective policies for problems with a constrained action space, where certain actions can only be performed a limited number of times. At the same time, action masking should also be used with care since enforcing overly strict heuristics can prevent the RL agent from learning optimal policies. Our study contributes to the operations management literature by outlining ways of incorporating human knowledge into machine learning methods, which also provides valuable guidance for managers opting to employ RL in real-world applications.

3 - Empowering Financial Decisions: Unveiling LLMs' Potential in Sentiment Analysis

Sorouralsadat Fatemi, University of Illinois at Chicago, Chicago, IL, United States, Yuheng Hu

Financial sentiment analysis is vital for informed decision-making in finance, leveraging Large Language Models (LLMs) for effective insights. While LLMs excel in various NLP tasks, their full potential in financial sentiment analysis remains untapped. We explore two strategies: in-context learning (GPT-3.5, GPT-4) and fine-tuning LLMs on finance data, focusing on smaller models (250M to 7B parameters) due to computational costs. Comparing with state-of-the-art results, we find that fine-tuned smaller LLMs match or surpass larger counterparts, even with less data. Zero-shot and one-shot performance also align well with fine-tuned models and benchmarks. Interestingly, increasing in-context learning shots doesn't significantly boost sentiment analysis performance in finance. This research highlights the promise of appropriately tailored LLMs for robust financial sentiment analysis, aiding decision-makers amidst market fluctuations.

4 - Developing a Machine Learning-Based Warning System for Predicting Graduation as Scheduled Based on University Student Course Performance

Hao Huang, Yuan Ze University, Taoyuan, Taiwan, Stevanus Calvin Widianto, Ray F. Lin

This study introduces an approach to constructing a warning system tailored to forecast students' likelihood of graduating as scheduled, predicated on their course performance. We implement random forest and artificial neural network algorithms as our prediction model. Through melding these advanced machine learning techniques with comprehensive academic data increases each semester, multiple models are created for different time periods of a student's progression. For instance, there is a model for predicting student's graduating as scheduled after the first semester, whereas another model is trained for the same prediction after the second semester with more course performance data. Overall, eight models are trained for the four-year period of a student's university progression. The numerical experiment is currently implemented with the students from the Department of Industrial Engineering to illustrate the potential of the warning system. This warning system endeavors to furnish timely interventions for students deemed at risk of academic setbacks, thereby potentially fortifying retention rates and nurturing scholastic achievement. In summation, this study engenders discourse on the potential of machine learning in education.

By forging a robust warning system, universities could be poised to identify and support students encountering academic hurdles, fostering a conducive environment for student success.

TB24

Summit - 344

Advanced Decision-Making and Optimization Techniques

Contributed Session

Chair: Yuan Dong, Temple University, philadelphia, PA, United States

1 - The Economic Impact of Environmental Product Declarations and Low-Carbon Certifications: A Case Analysis of South Korean Companies

Jeongmin Shin, Kyung Hee University, Yongin, Korea, Republic of, Jungwoo Shin

This study analyzes the performance of environmental performance declarations (EPDs) and low-carbon certification schemes to investigate whether labelling environmental performance alone can lead to significant results. The analysis of the performance of companies according to their certifications was conducted in Korea. We used a database containing information on 57,908 Korean companies and a list of 822 EPD certified companies and 282 low-carbon product certified companies published on the Environmental Technology Business One-Stop Service. To compensate for the lack of small and medium-sized enterprises (SMEs) data in the collected Korean company database, an additional company survey was conducted to build the final analysis database. The propensity score matching (PSM) methodology was used to compare certified and non-certified companies, and selection bias was adjusted to obtain direct results. The results showed that companies with Ecolabelling and low-carbon certification experienced a significant increase in sales between one and two years after certification. This suggests that environmental certification can contribute to improving corporate performance and increasing competitiveness in the marketplace.

2 - Improving Support Vector Ordinal Regression for Enhanced Performance in Outlier and Imbalanced Datasets

Stephan Jao, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Puli, Taiwan, Jing-Rung Yu, Chun-Yen Tsai, Donald Lien, Hung Pi-Hsia, Chi-Chun Chien

Support Vector Ordinal Regression (SVOR) is a classification method that classifies multiple classes with preference orders. Given a dataset with r ranks, the SVOR obtains r-1 parallel hyperplanes (classifiers) with r-1 thresholds in order to categorize the training data with the hinge loss function by maximizing the margins between ranks. Outliers and imbalanced data appear frequently in real-world datasets. The appearance of outliers in a dataset affects the magnitude of the classification error and hence influences the position of the hyperplanes, especially narrowing the range of ranks situated between hyperplanes, and resulting in lower classification accuracy. Furthermore, SVOR is weak at handling imbalanced data, especially when the classification outcome is affected by the majority class. Research shows resampling methods to address imbalance problems tend to suffer from information loss, overfitting, and increased model complexity. Meanwhile, cost sensitive learning maintains the completeness of the dataset. This study proposes an l_1-norm SVOR model with bounded loss functions. The training results are surveyed to evaluate the impact on outliers. In addition, we will also adjust the loss function by adding a higher penalty to the misclassification error in each corresponding rank. The proposed method is applied to an imbalanced dataset, which is composed of financial data of Taiwanese companies from 2009 to 2022 to predict financial distress. The data is composed of 3 ranks, with rank 1 being the majority rank. Comparisons highlight advantages of our study in classification accuracy and cost-effectiveness for dealing with outliers and imbalanced datasets.

3 - On the Approximability of the Yolk in the Spatial Model of Voting

Ran Hu, Rensselaer Polytechnic Institute, Troy, NY, United States, James Bailey

In the spatial model of voting, the yolk and LP (linear programming) yolk are important solution concepts for predicting outcomes for a committee of voters. McKelvey and Tovey showed that the LP yolk provides a lower bound approximation for the size of the yolk and there has been considerable debate on whether the LP yolk is a good approximation of the yolk. In this paper, we show that for an odd number of voters in a two-dimensional space that the yolk radius is at most twice the size of the LP yolk radius. However, we also show that (1) even in this setting, the LP yolk center can be arbitrarily far away from the yolk center (relative to the radius of the yolk) and (2) for all other settings (an even number of voters or in dimension $k \geq 3$) that the LP yolk can be arbitrarily small relative to the yolk. Thus, in general, the LP yolk can be an arbitrarily poor approximation of the yolk.

4 - The Impact of Investor Irrationality on Firm Value and Market Efficiency

Wei HU, School of Economics and Management, Tongji University, Shanghai, China, People's Republic of, Xiurui YANG

The short squeeze event involving GameStop (Ticker: GME) in 2021 was a landmark event that underscores the transformative power of market irrationality in shaping the financial markets. Retail investors launched a campaign through online forums and social media platforms to counter short-selling institutional investors by purchasing GameStop stock. Using a dataset from *Stocktwits*, a professional social media platform for stock discussions, we examine the causal effects of the GME short squeeze on the firm value and market efficiency for 367 SP500 listed stocks from October 2020 to April 2021. Firm value is assessed using abnormal returns, while market efficiency is proxied by the social media users' forecast accuracy (SMUFA). We utilized pre-trained language models to predict the sentiment of messages and employed graph embedding models to identify GME-related stocks. A difference-in-differences approach combined with propensity score matching reveals that GME-relevant stocks exhibited a significant decline in both abnormal return and SMUFA compared to GME-irrelevant stocks after the short squeeze. These findings suggest that market irrationality negatively impacts firm value and market efficiency. Our results, which successfully passed a series of tests and proved robust, offer significant theoretical contributions and practical implications.

5 - Media Optimization with Nbcuniversal One Platform Total Measurements

Hanxi Bao, NBCUniversal Media, LLC, New York, NY, United States

NBCUniversal launched One Platform Total Audience (OPTA) at beginning of 2024. The AI-driven planning and activation technology is the latest evolution within the company's One Platform offering and leverages new investments in automation and data to provide brands unduplicated reach through a single media buy across linear and streaming. OPTA allows advertisers to generate media plans across NBCU linear and digital media channels using more than 300 audience segments that connect audience attributes and outcomes. Together with OPTA, we established One Platform Total Measurement (OPTM) framework that delivers insights and outcome-based measurement across the entire funnel, from awareness to conversion. This presentation introduces the standardized measurement operating procedure and media planning optimization methods to find the optimal media mix at NBCUniversal to drive target outcomes.

6 - Product Pricing and Live Streaming E-commerce Frequency Decisions

Yuan Dong, Temple University, philadelphia, PA, United States, Guohou Shan, Subodha Kumar

Live streaming e-commerce (LSE) gradually becomes the future of shopping and creates hundreds of billions of markets. Influencers can demonstrate the product and interact with potential consumers in real-time on LSE platforms, reducing product uncertainty and raising the consumers' purchasing intention. In this research, we focus on understanding how online stores alter LSE frequency and pricing decisions to boost their revenue better. To better understand consumers' decisions, we also categorize consumers into product-oriented and influencer-oriented, aiming to understand the impact of consumer heterogeneity on store decisions. Our research makes both theoretical and practical contributions. On the one hand, we enrich the prior literature by focusing on exploring more nuanced decisions that a store can make regarding the LSE setting and adoption while considering customers' heterogeneity. On the other hand, our research findings can help online stores make better decisions and strategies as for leveraging LSE to boost their product sales and revenue.

TB25

Summit - 345

Provalis/AMPL

Invited Session

Technology Showcase

1 - Topic Modeling or Taxonomy Building: Benefits, Drawbacks, and Implementation Strategies

Normand Peladeau, Provalis Research, Montreal, QC, Canada

Topic modeling is a "quick and dirty" heuristic method to inductively extract topics from large text collections. While it holds significant potential, the results can sometimes be suboptimal. This presentation will explore various strategies to enhance topic modeling outcomes, including alternative algorithms, the application of large language models (LLMs) and word embeddings, and automatic word sense disambiguation techniques. We will also identify limitations of topic modeling and how building text analytics taxonomies can serve as a more precise alternative for measurement, capturing low-signal events, and hypothesis testing. Although developing such taxonomies is time-consuming, we will demonstrate, using WordStat text analytics software, how various techniques and strategies can yield more accurate results more efficiently and how these two approaches can complement each other.

2 - Python and AMPL: Build Optimization Applications Quickly with amplpy, Pandas, Streamlit - and AI

Bob Fourer, AMPL Optimization Inc., Evanston, IL, United States

Python and its vast ecosystem are great for data pre-processing, solution analysis, and visualization, but Python's design as a general-purpose programming language makes it less than ideal for expressing typical complex optimization problems. AMPL is a declarative language that is designed for describing problems and that integrates naturally with Python.

In this presentation, we'll survey a range of contexts where AMPL and Python work together to make optimization software simpler to use, faster to run, and easier to integrate with enterprise systems:
- Installing AMPL and solvers as Python packages anywhere
- Importing and exporting data efficiently from/to Python data structures, including Pandas and Polars dataframes
- Modeling and solving in Jupyter notebooks on Google Colab
- Deploying to the cloud quickly and easily with Pandas, Streamlit, and amplpy

You'll also see how generative AI technology is enabling a rapid development process for both AMPL and Python, reducing the time and effort to produce working applications that are ready for end-users.

TB26

Summit - 346

Digital Engagement and Innovation in Education and Social Networks

Contributed Session

Chair: Deanna Sanchez, 76202

1 - Effects of Digital Commitment Contracts on Engagement in Underprivileged Children's Education: Evidence from a Field Experiment

Swanand Deodhar, Indian Institute of Management Ahmedabad (IIMA), Ahmedabad, India, Neaketa Chawla, Ambrish Dongre, Harsh Parekh, Yana Chello, Neetu Singh, Prashast Srivastava

Students' continued engagement in education programs, particularly among students from underprivileged backgrounds, has emerged as a pertinent problem in enabling the benefits of education. To this end, digital technologies may act as a mitigating force. Our study ascertains how digital interventions may increase engagement in a primary education initiative. Drawing from prior work on commitment theories, we argue that a digitally-enabled *Commitment Contract* (CC) may aid in improving target groups' engagement. We report findings from a randomized field experiment conducted through a non-government organization (NGO), which works to improve primary education in underprivileged communities located in North India. As a treatment, we sent WhatsApp notifications, seeking the recipients' commitment to undergoing an assessment of basic skills in *Mathematics* and *Hindi* (a prominent Indian language) within 48 hours of receiving the message. Our dataset comprises recipients' responses to the CC and the subsequent phone call for assessment by the NGO. Our estimates, based on randomization, matching, and a post-experiment survey, reveal that, on average, CC improves the efficiency of the evaluation system as fewer calls are required to complete the same levels of assessment. Moreover, we find that those who do not commit after receiving the CC prompt exhibit lower engagement than those in the control group, indicating that the CC may help in segmenting recipients based on their commitment. Lastly, based on the average cost of an assessment observed before the experiment, we show significant economic implications of CC for a resource-constrained organization such as an NGO.

2 - Exploring the Impact of Extended Reality Training Systems on Nursing Education: A Comprehensive Physiological Study

Kamelia Sepanloo, Purdue University, West Lafayette, IN, United States, Young-Jun Son, Janine Hinton

This presentation outlines the methodological framework and objectives of pilot studies aimed at assessing the effectiveness of mixed reality (MR) and virtual reality (VR) training systems in nursing education, with a particular focus on collecting and analyzing physiological measurements. The study seeks to explore how VR and MR technologies influence various physiological responses among nursing students during their training sessions. Key physiological indicators, such as heart rate variability and activity energy expenditure, are closely monitored to gauge engagement levels, stress responses, learning effectiveness, emotional reactions, fatigue, and workload.

The research follows a structured methodology, beginning with the design of immersive VR and MR scenarios tailored specifically to nursing education. This is followed by the careful selection of appropriate physiological sensors for data collection and the execution of controlled training sessions involving both students and faculty. A diverse range of training tasks and scenarios are incorporated to provoke a spectrum of physiological responses, facilitating a thorough analysis.

Furthermore, in addition to physiological data, feedback on usability and comfort is solicited from participants to inform iterative enhancements to the VR/MR training systems. This emphasis on user experience underscores the importance of integrating physiological measurements into the evaluation process to comprehensively assess the impact of VR and MR training systems on nursing education.

Ultimately, this holistic approach aims to lay the groundwork for future advancements in immersive learning experiences, potentially leading to a transformative shift in the landscape of nursing education through innovative technological modalities.

3 - Alteryx Location Analytics in Higher Education

Deanna Sanchez, Nova Geographica LLC, Denton, TX, United States

Alteryx is a global low-code platform that enables quick data analytics, including spatial and demographic analysis capabilities. Location analytics in Alteryx encompasses many features, and this presentation examines the spatial relationships between student and university locations. We will explore geocoding addresses, distance calculations utilizing drivetime analysis, and using Census block group demographics to efficiently perform outreach to specific student populations.

4 - Not All Doom and Gloom: ChatGPT, Knowledge Diversity, and Style Similarity in Online Communities

Chan Gyu Lee, Seoul National University, Seoul, Korea, Republic of, Jeongsik Lee, Hyunwoo Park

Understanding the dynamics of user behavior in online collaborative communities (OCCs) is essential to leverage and maximize the potential of these communities. While OCCs heavily depend on users' voluntary involvement, concerns have been raised regarding potential declines in user engagement following the introduction of ChatGPT, an advanced large language model. This study examines the disruptive impact of ChatGPT on user behavior within Stack Overflow—an OCC that is highly prominent and arguably most severely impacted by the technology shock—focusing on knowledge coverage and community cohesion. For the investigation, we use the entropy and linguistic-style match (LSM) methods to measure content diversity and linguistic style similarity, respectively, within the community. We then apply a difference-in-differences analysis to assess the causal effect of ChatGPT on these measures. Despite an overall decrease in the volume of user-generated posts following the introduction of ChatGPT, our empirical findings reveal a positive impact on knowledge coverage, indicating that ChatGPT helps enhance content diversity within the community. Additionally, the results indicate a trend of linguistic style homogenization, suggesting a more cohesive community environment potentially enabled by ChatGPT. Our study provides some early empirical evidence that, contrary to initial concerns about the introduction of ChatGPT threatening the role of OCCs, ChatGPT may have instead enriched and diversified community knowledge, fostering a more cohesive and collaborative environment for these communities.

TB27

Summit - 347

Product and Process Innovation in Healthcare

Invited Session

New Product Development

Chair: Zhili Tian, Univ. of Houston, Houston, TX, United States

1 - Optimal Enrollment In New Drug Development With Learning of Drug' Efficacy for Clinical Trials

Zhili Tian, Univ. of Houston, Houston, TX, United States

The cost for developing a new drug ranged from $1 billion to more than $2 billion between 2010 and 2019. The efficacy of the candidate drug, patient enrollment, and the market exclusivity period (MEP) are uncertain. Moreover, slow enrollment leads to increased costs, canceled clinical trials, and lost potential revenue. Many firms hope to detect efficacy versus futility of the candidate drug early to save development costs in group-sequential-design trials. We developed a continuous-time dynamic programming model with learning of a drug's efficacy and MEP. We update a drug's efficacy by Bayes' rules and prove the value function and optimal properties using the stochastic order and the likelihood-ratio order of distribution functions. Our optimal policies can assist firms in developing optimal enrollment policies in their clinical trials. Firms can use the properties of the value function to select late-stage clinical trials for their drug-development project portfolios, use our optimal policy to guide patient recruitment in their clinical trials considering competition from other drugs in the marketplace, and use our model in simulation to select their trial design parameters.

2 - Digital Inclusion: Can Platform Affordances Mitigate Online Harassment?

Eujean Byun, University of California Merced, Merced, CA, United States, Lisa Yeo, Marie Yeh

Online harassment is defined as repeated malicious internet behaviors intended to upset or disturb victims, often conveying explicit or implicit threats. While anyone can be a victim, there is ample evidence that younger and marginalized people (i.e., vulnerable groups) experience online harassment more frequently and of greater severity; designing interventions is a critical tool in supporting justice, equity, diversity, and digital inclusion. This study examines strategies to reduce online harm by promoting bystander intervention; we further explore if personality traits play a role in moderating bystander intervention behaviors and their attitudes. Using the Truman Platform (simulated social media engine), we manipulate intervention design affordances of a social media platform and observe the impact on user intervention behaviors and their attitudes. We recruited participants via Prolific from February to April 2024, with 137 individuals completing the experiment. Participants were exposed to varying levels of harassment severity, reporting effort, dissent visibility, and rules reminders in $2\times3\times2\times2$ full factorial treatment groups and were randomly assigned to a treatment group. Participants completed a pre-experiment survey assessing select personality traits, engaged with our simulated social media platform over two days, and then completed a survey about their experience along with social behavior, and were debriefed about the deceptive nature of the platform. This study will inform better design to support and encourage bystanders to intervene, while also anticipating extending the work to address how to avoid the weaponization of these same systems to cause further harm to marginalized individuals and groups.

3 - Plastic Recycling Fueled with Blockchain-Driven Tokenization: ESG Optimization and Implications

Jim (Junmin) Shi, New Jersey Institute of Technology, Newark, NJ, United States, Aichih Chang, Nesreen El-Rayes, Fuqin Zhou

The plastic waste crisis has been witnessed to be exacerbating nowadays. To address the burgeoning challenge, one innovative solution is to incentivize recycling via leveraging tokenization thanks to Blockchain technology (BCT). In this study, we device an optimization model considering the tokenization to examine the environmental and social governance (ESG) performance. In particular, we investigate the optimal decisions of tokenization rewards and the government enforcement as well as their interplay on the ecosystem. We demonstrate the potential for significant ESG improvements by managing enforcement and tokenization jointly. Extensive numerical experiments and sensitivity analysis are performed to provide rich insights. For example, the value of tokenization in terms of ESG is visualized.

4 - Chinese EV Manufacturers: Integrating Industry 4.0 Technologies with Lean and Sustainable Manufacturing Practices to Elevate Operational and Innovation Performance, leading to sustainable production

Muhammad Adnan Zahid Chudhery, School of Management, University of Science and Technology of China, Hefei, China, China, People's Republic of, Sarah Safdar, Huafei Wei

In 2023, Chinese electric vehicle (EV) sales reached 9.49 million units. This included 2.80 million plug-in hybrids and 6.68 million pure battery-powered vehicles. Some Chinese EV manufacturers, such as BYD, are now competing with international players like Tesla to gain market dominance. Chinese auto manufacturers are currently facing the challenge of oversupply, leading to a fiercely competitive environment and a price war, resulting in reduced profitability. Experts predict that many smaller companies will not survive this situation. In this scenario, integrating Industry 4.0 with lean and sustainable manufacturing is crucial for these companies to achieve operational and innovative performance for sustainable production. Therefore, we conducted this study in Chinese EV manufacturing companies and collected data from knowledge employees and their direct managers. We argued that lean and sustainable manufacturing practices could help Chinese EV manufacturers elevate their operational and innovation performance, resulting in sustainable production using Industry 4.0 technologies. The study has significant theoretical implications for researchers and practical implications for government environmental authorities, policymakers, and EV manufacturing companies.

TB28

Summit - 348

Novel Analytics and Machine Learning Approaches with Manufacturing Applications

Contributed Session

Chair: Thiraj Wegala, Auburn University

1 - Machine learning-assisted ultrasonic testing and its application in additive manufacturing

Thiraj Wegala, Auburn University, Auburn, AL, United States, Rongxuan Wang

Porosities are considered one of the most important defects in additive manufacturing as they directly influence the part's strength. The porosity level assessment is usually ex-situ and relies on expensive equipment such as XCT. Ultrasonic imaging has been used for internal metal defect detection but can only be resolved with millimeter-level accuracy, not enough for micrometer-level porosity detection in laser powder bed fusion. The main reason is the physics-driven model is only able to track strong and simple echoes from relatively large defects. In this work, a data-driven model is trained to demodulate weak and complex diffraction signals. When shooting a group of pulses with different phases and wavelengths into the part, some sound waves will be trapped or deflected by the porosities. Two ultrasonic microphones are used to collect those sound waves that being influenced by the porosities. They are mounted at different locations and angles to provide spatial information. The amplitude signal of the microphones is used as the input of the data-driven model, and the output is a porosity level map on the top surface of the part. With this information, remelting can be performed to release the porosities at the location that does not meet the requirements.

2 - Multiphysics-Informed Machine Learning using Sequential Learning for Surface Mount Technology

Abdelrahman Farrag, State University of New York At Binghamton, Binghamton, NY, United States, Jaewoo Kim, Daehan Won, Yu Jin

The electronics industry's continual drive towards miniaturization has posed unique challenges for Surface Mount Technology (SMT), especially in the assembly of printed circuit boards (PCBs). This paper introduces the ThermoDynaSMT framework to tackle the critical issue of predicting the final position of small-sized passive components, Self-alignment behavior, post the Soldering Reflow Process (SRP). The unpredictability of component placement arises due to a blend of factors including the surface tension of the solder paste, capillary effects, and the differential surface energy between the solder mask and copper pad. This behavior is modeled as a multi-physics problem considering the heat transfer principles governing the SRP and the mechanics influencing the movement of the components. ThermoDynaSMT integrates 2D Physics-Informed Neural Network (PINN) for predicting temperature distribution across the PCB, and a hybrid machine learning model for determining the final location of components with high precision. The ThermoDynaSMT framework combines real-time thermal profile adjustments with advanced predictive capabilities, achieving component placement precision within 10 micrometers. This precision is essential for maintaining the integrity and functionality of the assembled PCBs, ensuring adherence to Industry 4.0 standards. The model addresses limitations found in traditional methods, which typically rely on experimental data and trial-and-error adjustments, by employing sequential learning between its sub-networks to enhance stability and accuracy. Additionally, adaptive loss term weights are introduced to optimize the multiphysics modeling process in the PINN model, further improving the predictive control and enhancement of the manufacturing process in response to the evolving technological demands in the electronics assembly industry.

3 - A Mathematical Model to Optimize Capacity of a Complex Semiconductor Fabrication Plant

Jinyoun Lee, Samsung Electronics, Hwaseong-si, Korea, Republic of, Sungha Park, Seungho Lee, Euiseok Kum

In semiconductor FAB construction, substantial investment is required, making accurate assessment of FAB capacity crucial. This involves considering the complex nature of semiconductor FABs, including not only equipment perspective but also the analysis from a chamber-level perspective. Moreover, considering operational aspects such as complex product mix, priorities, and survival rates due to factors like destructive testing is crucial for accurate decision-making. This study proposes a mathematical model that considers chamber-level arrangement information and operational aspects of products, facilitating the derivation of optimal wafer production and capacity operation strategies. Additionally, it aids in identifying and analyzing bottleneck equipment or chambers during operations, thereby optimizing FAB capacity operation strategies in line with investments.

4 - Developing Integrative Production-Delivery System using Mobile Additive Manufacturing and Truck-Drone Delivery

Junfeng Ma, Mississippi State University, Mississippi State, MS, United States, Omid Abdolazimi

Recent advancements have seen the implementation of mobile additive manufacturing (MAM), which involves 3D printing technologies in moving vehicles, garnering considerable interest from both academic circles and industry sectors. Despite the potential benefits each technology offers to supply chain management, research integrating MAM with truck-drone delivery systems for daily operations is lacking. This study addresses this shortfall by combining these innovative technologies while also taking into account the preferred delivery time window of customers and the optimal sequence for printing. A mixed integer linear programming model is developed to facilitate this novel approach. To solve the model, both an exact method using CPLEX and a heuristic method via the Lagrangian relaxation algorithm are employed and evaluated. Computational experiments confirm the efficacy of the model and methods across small, medium, and large-scale problems, with the Lagrangian relaxation algorithm consistently outperforming in terms of optimality gap and CPU time. Furthermore, the model is applied in a real case study utilizing the Lagrangian approach. Findings reveal that the maximum drone flight range and the complexity of the printing tasks notably influence overall costs and meet customer delivery timelines. This system not only meets customer needs in a timely and cost-efficient manner but also enhances supply chain resilience, enabling businesses to boost service quality robustness and lessen reliance on inventory by producing spare parts on-demand through MAM technology.

5 - Different Batch Composition for Additive Manufacturing Machines in an Unrelated Parallel Machine Scheduling Problem

Fatemeh Pourdehghan, University of Missouri, Columbia, MO, United States, Ayhum Ishaq Asfour, Omid Shahvari, Sharan Srinivas

In the next decade, Additive Manufacturing (AM) will revolutionize component production with complex geometries, enabling on-demand creation at or near end- use locations. This shift reduces lead times and inventory, cuts transport distances, and enhances sustainability by lowering resource use and emissions. By using this technology, the manufacturer can quickly change for a revised version of the product which has less material waste during production. Where a degree of customization is required, particularly in additive manufacturing, considering an unrelated parallel machine layout is beneficial due to its structured production process to optimize efficiency. Our study focuses on batch processing in additive manufacturing on unrelated parallel machines, aiming to minimize total cost, including production cost, tardiness cost and rejection cost. To reflect real-world industry requirements, variable machine availability, dynamic job release times, different machine capability and eligibility, multiple raw materials, batch capacity constraints, sequence-dependent setup time, and customer prioritization are considered. Unlike many prior works, this study considers different batch compositions for machines. While the batch capacity depends on the raw material, the batch compositions depend on the required raw material, dimension, and volume of the jobs, as well as the necessary spacing between jobs in a batch. A mixed-integer linear programming model is proposed for small-size problems. Given the problem's NP-hard nature, we also introduce Simulated Annealing algorithm to tackle large size problems. The results demonstrate the algorithm's high level of effectiveness.

6 - A Hybrid Genetic Algorithm for Topology Optimization of Frame Structures for Additive Manufacturing

Oguz Toragay, Lawrence Technological University, Southfield, MI, United States, Nikan Mehrara, Daniel Silva, Alexander Vinel

In this ongoing research we consider the design of lightweight additively manufactured planar frame structures as a nonlinear, nonconvex optimization problem for which we propose a novel hybrid Genetic Algorithm (GA). Our motivation in proposing the metaheuristic is that the large-scale instances of frame structures cannot be tackled using existing exact methods from the literature. The proposed method is a hybrid GA, in that it consists of solving a nonlinear optimization problem to calculate the fitness values of the individuals in the population, in addition to other stochastic behaviors of classical GA. We also propose a novel encoding for the candidate solutions as the individuals in the GA's population, based on the paths that can be identified in the ground structure. Unlike traditional binary encoding, path-based encoding considers the mechanical behaviors of the structure and makes GA applicable in the discrete structure's topology optimization problem. Accordingly, we modified operators for the hybrid GA to handle the path-based encoding. Our numerical results showed that although the proposed hybrid GA is able to find the optimal solution for small-scale problems, the accuracy and consistency of the results suffer from the curse of dimensionality. We conclude that the proposed path-based encoding is a capable alternative for the traditional binary encoding that has been widely used for combinatorial optimization problems in the literature.

TB29

Summit - 420

Topics in Integer Programming and Combinatorial Optimization 2

Invited Session

OPT: Integer and Discrete Optimization

Chair: Luze Xu, University of California, Davis, Davis, CA, 95616, United States

Co-Chair: Joseph Paat, University of British Columbia, Sauder School of Business, Vancouver, BC, Canada

1 - Breaking the quadratic gap for strongly polynomial solvers to combinatorial linear programs

Bento Natura, Georgia Institute of Technology, Atlanta, GA, United States

Recent years have seen tremendous progress in high-accuracy solvers for Maximum Flow, Minimum-Cost Flow and general Linear Programs (LP). Progress on strongly polynomial solvers for combinatorial LP on the other hand has stalled. For combinatorial LP beyond directed graphs this gap between exact and high-accuracy solvers is currently quadratic. We finally break the quadratic gap and design a strongly polynomial interior-point-method for combinatorial LP, which reduces the gap to only a linear factor.

2 - Fast algorithm design for structured linear programming

Sally Dong, MIT, Cambridge, MA, United States

An extremely fruitful line of algorithms research over the past decade has been the application of interior point methods alongside data structure design to classical problems in combinatorial optimization. In this talk, we consider linear programs of the form min <c,x> subjected to Ax = b, x >= 0, where the constraint matrix A has suitable structural properties. We present a general framework for solving these structured linear programs that ties together interior point methods and tools across theoretical computer science including graph decomposition, sampling, dynamic algorithms, and numerical linear algebra. Our framework in turn yields the fastest-known algorithms for min-cost flow and k-commodity flow on planar graphs, and for min-cost flow and general linear programs on graphs with bounded treewidth.

Based on joint work with Yu Gao, Gramoz Goranci, Yin Tat Lee, Lawrence Li, Richard Peng, Sushant Sachdeva, and Guanghao Ye.

3 - Sample Complexity of Algorithm Selection and its Applications to Branch-and-cut

Sammy Khalife, Johns Hopkins University, Baltimore, MD, United States

Can we learn efficiently good cutting planes to solve integer programs faster? What do we mean by efficient learning, and what is a good cutting plane? In this talk, I will delve into these questions, building upon recent work in data-driven algorithm design. This paradigm uses statistical and machine learning techniques to select from a class of algorithms for a computational problem an algorithm that has the best expected performance with respect to some (unknown) distribution on the instances of the problem. In this context, instead of selecting a single algorithm that has the best performance, we allow the possibility of selecting an algorithm based on the instance to be solved. In particular, given a representative sample of instances, we learn a neural network that maps an instance of the problem to the most appropriate algorithm for that instance. By formalizing this idea, I will present rigorous sample complexity bounds for this learning problem, in the spirit of recent work in data-driven algorithm design. We then apply this approach to the problem of making good decisions in the branch-and-cut framework for mixed-integer optimization (e.g., which cut to add?). In other words, the neural network will take as input a mixed-integer optimization instance and output a decision that will result in a small branch-and-cut tree for that instance. Our computational results provide evidence that our particular way of using neural networks for cut selection can make a significant impact in reducing branch-and-cut tree sizes, compared to previous data-driven approaches.

4 - Improving Strong-branching Decisions With Additional Information

Prachi Shah, ISyE, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey

Strong-branching uses only the local gains in the linear programming (LP) bounds to guide the branching decision. In this work, we aim to refine it by incorporating information from primal bounds and the problem structure.

5 - Decomposition-Based Disjunctive Cutting Planes

Aleksandr Kazachkov, University of Florida, Gainesville, FL, United States, Ricardo Fukasawa

Many integer programs have either explicitly user-defined decomposition structure, such as in multistage stochastic integer programming, or naturally-occurring decompositions that can be automatically detected. Each part of the decomposition typically involves a small fraction of the larger problem, which allows for expensive computational techniques on the resulting subproblem that would be impractical for the entire problem. We investigate the generation of sparse disjunction-based cutting planes using the decomposition structure of instances.

TB30

Summit - 421

Submodularity and Conic Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Georgina Hall, INSEAD, Fontainebleau, France

1 - Distributionally Robust Optimization Through the Lens of Submodularity

Karthik Natarajan, Singapore University of Technology and Design, Singapore, Singapore, Arjun Ramachandra, Divya Padmanabhan

Distributionally robust optimization is used to solve decision making problems under adversarial uncertainty where the distribution of the uncertainty is itself ambiguous. In this paper, we identify a class of these instances that is solvable in polynomial time by viewing it through the lens of submodularity. This adds to the list of known polynomial time solvable instances of the multimarginal optimal transport problem and the generalized moment problem by bridging ideas from convexity in continuous optimization to submodularity in discrete optimization. In turn, we show that a class of distributionally robust optimization problems with discrete random variables is solvable in polynomial time.

2 - Sequential Competitive Facility Location: Exact and Approximate Algorithms

Ruiwei Jiang, University of Michigan, Ann Arbor, MI, United States, Mingyao Qi, Siqian Shen

We study a competitive facility location problem (CFLP), where two firms sequentially open new facilities within their budgets, in order to maximize their market shares of demand that follows a probabilistic choice model. This process is a Stackelberg game and admits a bilevel mixed-integer nonlinear program (MINLP) formulation. We derive an equivalent, single-level MINLP reformulation and exploit the problem structures to derive two valid inequalities based on submodularity and concave overestimation, respectively. We use the two valid inequalities in a branch-and-cut algorithm to find globally optimal solutions. Then, we propose an approximation algorithm to find good-quality solutions with a constant approximation guarantee.

3 - Sequential Submodular Maximization

Arash Asadpour, Zicklin School of Business, New York, NY, United States, Rad Niazadeh, Amin Saberi, Ali Shameli

We study a submodular maximization problem motivated by applications in online retail. A platform displays a list of products to a user in response to a search query. The user inspects the first k items in the list for a k chosen at random from a given distribution and decides whether to purchase an item from that set based on a choice model. The goal of the platform is to maximize the engagement of the shopper defined as the probability of purchase. This problem gives rise to a less-studied variation of submodular maximization, in which we are asked to choose an *ordering* of a set of elements to maximize a linear combination of different submodular functions. First, using a reduction to maximizing submodular functions over matroids, we give an optimal approximation for this problem. We then consider a variant in which the platform cares not only about user engagement, but also about diversification across various groups of users—that is, guaranteeing a certain probability of purchase in each group. We characterize the polytope of feasible solutions and give a bicriteria approximation for this problem by rounding an approximate solution of a linear-programming (LP) relaxation. For rounding, we rely on our reduction and the particular rounding techniques for matroid polytopes. For the special case of coverage functions—which is practically relevant in online retail—we propose an alternative LP relaxation and a simpler randomized rounding for the problem, yielding an optimal bicriteria approximation algorithm.

4 - Sum of Squares Submodularity

Georgina Hall, INSEAD, Fontainebleau, France

Submodularity is a key property of set-valued functions, appearing in many different areas such as game theory, machine learning, and optimization. In this talk, we consider set-valued functions with a multilinear extension of degree less than or equal to some k. For these functions, it is known that testing whether they are submodular is hard for k greater or equal to 4. We thus introduce a sufficient condition for submodularity based on sum of squares polynomials, which we refer to as sum of squares (sos) submodularity. We investigate the gap between submodularity and sos submodularity and show how to use semidefinite programming to optimize over this set. In particular, we use the techniques proposed to learn a submodular function from data, among other applications.

TB31

Summit - 422

Advances in Non-Smooth Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Tianjiao Li, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Zhe Zhang, Georgia Tech, Atlanta, GA, United States

Co-Chair: Liwei Jiang, Georgia Institute of Technology, Atlanta, GA, United States

1 - On Reaching Local Minima of Nonconvex Nonsmooth Functions

Swati Padmanabhan, Massachusetts Institute of Technology, Cambridge, MA, United States

For Lipschitz functions, it is known that first-order algorithms converge to local minima asymptotically (due to a result by Davis, Druvyatskiy, and Jiang).
The natural next question then is whether these algorithms converge to local minima ``fast''. We prove a surprising negative result for this question. Specifically, we show that obtaining any non-trivial convergence rate in terms of local function decrease is impossible even for strict functions.
In particular, we prove that even under suitable regularity assumptions and the absence of non-strict saddles,
any algorithm whatsoever based on local queries cannot guarantee getting stuck at points at which there is significant local decrease,
unless the number of iterations grows exponentially with the dimension. In fact, this statement holds even under the supposedly easier case for which all approximate-stationary points below some constant (sub)gradient norm, are in fact global minima (which trivially precludes the existence of non-strict saddles).

This is based on joint work with Guy Kornowski and Ohad Shamir.

2 - Variational Theory and Algorithms for a Class of Asymptotically Approachable Nonconvex Problems

Ying Cui, University of California, Berkeley, Berkeley, CA, United States

In this talk, motivated by the inverse optimal value based non- linear optimization problems, we introduce a class of composite nonconvex functions, where the outer function is the sum of univariate extended-real-valued convex functions and the inner function is the limit of difference-of-convex functions. A notable feature of this class is that the inner function can be merely lower semicontinuous instead of continuously differentiable. It covers a range of important yet challenging applications, including the composite value functions of nonlinear programs and the value at risk constraints for continuous random vectors. We propose an asymptotic decomposition of the composite function that guarantees epi-convergence to the original function, leading to necessary optimality conditions for the corresponding minimization problem. The proposed decomposition also enables us to design a numerical algorithm to solve the composite minimization problem such that any accumulation point of the generated sequence, if exists, satisfies the newly introduced optimality conditions. Our derived results expand on the study of so-called amenable functions introduced by Poliquin and Rockafellar in 1992, which are compositions of convex functions with smooth maps, and the prox-linear methods for their minimization.

3 - A Superlinearly Convergent First-Order Method for Nonsmooth Optimization

Vasilis Charisopoulos, University of Chicago, Chicago, IL, United States, Damek Davis

Nonsmooth optimization problems appear throughout machine learning and signal processing. However, standard first-order methods for nonsmooth optimization can be slow for "poorly conditioned" problems. In this talk, I will present a locally accelerated first-order method that is less sensitive to conditioning and achieves superlinear (i.e., double-exponential) convergence near solutions for a broad family of problems. The algorithm is inspired by Newton's method for solving nonlinear equations.

4 - Linearly Convergent Algorithms for Nonsmooth Problems with Smooth Pieces

Jimmy Zhang, IE, Purdue University, West Lafayette, IN, United States, Sra Suvrit

Non-smoothness is a major bottleneck to efficient optimization. In the absence of smoothness, the theoretic convergence guarantee drops from linear to sublinear rates for convex programs, and becomes orders of magnitude worse for nonconvex programs. This huge gap is shown to be unimprovable for some worst-case non-smooth instances. Here we focus on some mild non-smoothness: the piecewise smooth (p.w.s) function whose domain can be partitioned into subsets such that the function restricted to each subset is smooth. The p.w.s function covers the relu function in neural network, the l1 penalty in sparse optimization, and the min-max saddle point problem in robust optimization as special cases. We present some **globally** linear convergent methods for optimizing the convex p.w.s smooth function satisfying the quadratic growth condition, and, as a corollary, we improve the iteration complexity for solving the weakly convex p.w.s problem by orders of magnitude. Importantly, our method does not require any knowledge about individual smooth pieces, thus it is applicable even to general non-smooth program exhibiting some local p.w.s structure.

TB32

Summit - 423

Efficient Methods for Large-scale Optimization Using Inexactness II

Invited Session

OPT: Optimization Under Uncertainty

Chair: Robert Baraldi, Sandia National Laboratories, Seattle, WA, 87123, United States

Co-Chair: Johannes Milz, Georgia Institute of Technology, Atlanta, GA, United States

1 - Memory-Efficient Dynamic Optimization using Inexact Hessians and Randomized Sketching

Radoslav Vuchkov, Sandia National Laboratories, Albuquerque, NM, United States

Finding solutions to dynamic optimization problems can be hampered by the sheer amount of memory needed to store the state trajectory, which is crucial for evaluating both the objective function and its derivatives. Recently, a trust-region method proposed by [R. Muthukumar et al., SIAM Journal on Optimization 31(2), pp. 1242–1275 (2021)] addressed this issue by using randomized sketching to compress the state trajectory, leading to inexact gradient calculations. The method provided convergence guarantees thanks to its adaptive rank learning criteria. Our work builds upon that approach by incorporating secant Hessian approximations. However, the inherent randomness of sketching can render traditional secant update formulas ineffective in producing accurate Hessian approximations. This is because the difference between two gradients, calculated from distinct sketches, might lack consistency. To circumvent this, we introduce a sketched approximation of the

Hessian itself, eliminating the need for gradient difference calculations. The effectiveness of our approach is showcased through numerical experiments.

2 - Derivative-free stochastic optimization via regression-based objective function models

Johannes Milz, Georgia Institute of Technology, Atlanta, GA, United States, Anton Kleywegt, Siva Ramani

We develop algorithmic approaches for optimization problems in which only noisy function evaluations are available. Motivated by applications in which such evaluations are costly, our framework is specifically designed to use both previous and new evaluations. We construct local objective function models, using evaluations from carefully selected experimental design points around the current iterate. We develop model-based algorithms such as trust region methods and cubic regularization methods that put function evaluations to best use while taking the model error into account in step selection. Under mild assumptions, we establish the convergence of our algorithms. Numerical illustrations and comparisons are presented.

3 - Approximate Internal Models for Power System Control Design in Extreme Conditions

Sergio A. Dorado-Rojas, Georgia Institute of Technology, Atlanta, GA, United States, John Cortés-Romero, Daniel Molzahn

The internal model principle has been one of the most effective yet underlooked tools for synthesizing control systems. In essence, the principle states that if a model of the exogenous signal is embedded in the control algorithm, the main performance objectives, such as disturbance rejection and reference tracking, are attained. However, the requirement for an exact model might be too stringent in some applications, such as modern power systems. This talk will revisit the internal model principle and show application examples in power systems where approximate internal models emerge as promising solutions for satisfactory performance in adverse or extreme conditions.

TB33

Summit - 424

Mitigating Distributional Ambiguity in Stochastic Programming

Invited Session

OPT: Optimization Under Uncertainty

Chair: Shixuan Zhang, Texas A&M University, College Station, TX, United States

1 - Value of Data for Distributionally Robust Optimization with Phi-Divergences

Jangho Park, Sungkyunkwan University, Suwon-si, Korea, Republic of, David Love, Guzin Bayraksan

In data-driven distributionally robust optimization, not all data has the same value on the optimal decision. This paper studies the value of additional data for distributionally robust optimization with phi-divergences. Additional data provides more information about the collected events. As a result, the relative information of the collected events increases compared to the uncollected events, resulting in a new optimization problem. We discuss whether the optimal value increases or decreases if one particular data is acquired and propose simple ways to estimate this value change before solving a new optimization problem. For situations where we do not know what value to be selected next, we discuss simple estimators for an average value change by considering all observed scenarios. This estimator helps to decide whether obtaining additional data is worthwhile or not before solving a new optimization problem. A necessary implication of this paper is that proposed estimators form the lower bounds of corresponding true values.

2 - Synthesizing Interactive Optimization and Expert Elicitation for Stochastic Programming Problems

John Nichols, Rensselaer Polytechnic Institute, Troy, NY, United States, Jennifer Pazour, Sandipan Mishra, John Mitchell

A way to better obtain informed estimates for input parameters required for stochastic programming problems is to combine interactive optimization and expert elicitation protocols. This research creates a framework that includes two connected optimization models that allow one to iteratively elicit a human expert for more information on stochastic input parameters. Preliminary results show that such an approach can reduce human elicitation efforts while achieving high design performance from the optimization model.

3 - Moment Relaxations for Wasserstein Distributionally Robust Optimization

Shixuan Zhang, Texas A&M University, College Station, TX, United States, Suhan Zhong

To mitigate in the distributional ambiguity in stochastic optimization problems, a popular approach is to consider a Wasserstein distance-based distributionally robust optimization (WDRO) counterpart. Beyond out-of-sample performance guarantee, such WDRO model allows adjustable in-sample conservatism that ensures the optimal decisions are asymptotically consistent. However, similar to robust optimization counterparts, WDRO often takes away convexity of the model and leaves an NP-hard problem associated with finding the worst-case expected value. Our work begins with the investigation of convex relaxations of these challenging WDRO problems where the objective function and constraints are polynomial. In this case, tight relaxations can be derived based on the moment information of the polynomials, and we show a simple condition that guarantees the same asymptotic order out-of-sample guarantee and in-sample adjustable conservatism as the original WDRO. This suggests that one may entirely use the moment relaxation as a surrogate for WDRO without asymptotic gaps. We provide numerical experiment results to illustrate these findings.

4 - Distributionally robust optimization-based infrastructure system maintenance planning

Zhengsong Lu, University of Pittsburgh, Pittsburgh, PA, United States, Bo Zeng

We develop a distributionally robust optimization model for the maintenance problem of an infrastructure system.

TB34

Summit - 425

Optimization and Modeling in Industrial Applications

Contributed Session

Chair: Mojtaba Salarpour, Texas A&M University-Commerce, Commerce, TX, United States

1 - Optimizing Project Schedule and Cost for Construction of a Floating, Production, Storage, and Off-loading (FPSO) Unit

Lu Gao, University of Houston, Houston, TX, United States, Bruno Camargo

A floating, production, storage, and offloading (FPSO) is a floating vessel. The first was built in 1977 and used for the production and storage of hydrocarbon. Its use is cost-effective in particularly at deep water locations and requires low maintenance costs. Due to the volatility of the Oil&Gas markets, it has become increasingly important for organizations to develop innovative business strategies and cost-saving measures to counter the downturns of the industry. To answer this dilemma, FPSOs have emerged as an effective, efficient, and economical means to capitalize on oil fields around the world. FPSOs provide the same functionality as fixed platforms with the flexibility of relocating as needed in response to market conditions. In this project, we evaluated an FPSO erection project with the intent of identifying the optimal solution that will generate the shortest construction schedule and lowest possible project cost. Therefore, we developed an algorithm to detect the minimum cost, taking into consideration the fabrication (welding, NDT and Hydrotest) of piping streams, based on workers' salary, availability, and activities sequence. We also incorporated a daily bonus payment to incentive the fabrication company going ahead with the schedule. The results demonstrated that depending on the bonus value and workers' cost, the best-optimized solution will modify the activity sequence. On a large scale, this algorithm could be used to develop and solve real construction issues in the Oil&Gas industry, bringing direct revenue to the companies and improving the schedule with through an incentive bonus.

2 - Optimization of Prime Power and DRA Consumption in Refined Product Pipeline

Ling Zu, Colonial Pipeline, Alpharetta, GA, United States, Anthony Leo, Min Wang

In today's pipeline operations scenario, improving efficiency and lowering operational costs remains a top priority for industry stakeholders. This study investigates the optimization model to determine the optimal power utilization and drag reduction agent (DRA) usage in pipelines. Understanding how economic factors and operational performance interact, this study aims to clarify approaches that balance the need to save costs with the necessity of maintaining fluid integrity and continuous throughput. By leveraging mathematical modeling, optimization algorithms and cost-benefit analysis, it seeks to develop strategic solutions that balance economic considerations with operational efficiency. Through this interdisciplinary approach, the research aims to provide actionable insights that can inform decision-making processes and drive improvements in pipeline management practices.

3 - Equilibria in Interdependent Natural-Gas and Power Markets: An Analytical Approach

Amir Mousavian, Clarkson University, Potsdam, NY, United States, Beheshteh Raouf, Antonio Conejo

Natural-gas and power systems and their corresponding markets have evolved over time rather independently. However, both systems are increasingly interdependent since combined cycle gas turbines (that use natural gas to produce electricity) increasingly couple them together. Therefore, suitable analysis techniques are most needed to comprehend the consequences on market outcomes of an increasing level of integration of both systems. There is a vast literature on integrated natural-gas and power markets assuming that the two markets are operated centrally by a single operator. This assumption is often not true in the real world, which necessities developing models for these interdependent yet independent markets. In this vein, this paper addresses this gap in the literature and provides analytical Nash-Cournot equilibrium models to represent the joint operation of natural-gas and power markets with the assumption that the market participants in each market make their own decisions independently seeking maximum profits, as often is the case in the real world. We develop an analytical equilibrium model and apply the Karush-Kuhn-Tucker (KKT) approach to obtain Nash-Cournot equilibria for the interdependent natural-gas and power markets. We use a double-duopoly case to study the interaction of both markets and to derive insightful analytical results. Moreover, we derive closed-form analytical expressions for spot-market equilibria in both natural-gas and power markets, which are relevant and of practical significance for decision makers. We complement the duopoly-duopoly study with a detailed sensitivity analysis.

4 - Mastering Global and Local Variance in Mixed-Model Assemblies: Why the Variable Takt Outperforms the Fixed Takt

Philipp Pithan, WHU - Otto Beisheim School of Management, Vallendar, Germany, Arnd Huchzermeier

As demands on assembly lines rapidly increase due to product individualization, geopolitical risk hedging, sustainability requirements, and the introduction of new technologies, the fixed takt is severely challenged. To cope with variance in processing times across products and product variants, assembly line balancing typically deploys i) feeding lines, ii) floaters, iii) parallel stations, or iv) circular assembly for the efficient manufacturing on the same line. Product complexity is ultimately traded off against line inefficiencies.

We develop a general assembly line balancing model using the variable takt that optimally utilizes manufacturing resources and allocates work to them. Under very general assumptions, we show that a variable takt strictly outperforms the fixed takt with regard to line utilization and profitability when either i) a maximum takt or ii) a weighted-average takt time (WATT) is adopted. Furthermore, an increase in product individualization can be handled much more efficiently for the variable takt than for the fixed takt. When demand or mix uncertainty or seasonality increases, the fixed takt performs arbitrarily worse.

Assembly lines can be fitted to realize the shortest delivery lead time while achieving a more resource-efficient production of highly customized products, allowing for further profit growth. The choice of a fixed takt system not only constrains companies' customer centricity but also poses stress on the workforce. In a VUCA world, industries need to shift to a variable launching rate of products to eliminate process inefficiencies, worker stress, and constraints on product individualization, respectively (technology) innovation.

5 - Global Competition in Semiconductor Manufacturing: A Two-Stage Stochastic Supply Chain Network Model Approach

Mojtaba Salarpour, Texas A&M University-Commerce, Commerce, TX, United States, Kian Zeinolabedinzadeh

Semiconductors are essential to modern technology and nowadays they are widely used in many different fields, including computing, communications, medical technologies, and defense capabilities. The recent disruptions in the semiconductor supply chains, exacerbated by the COVID-19 pandemic, have exposed significant vulnerabilities in their global supply chain and have intensified the competition for leadership in semiconductor production globally so that the world's major powers consider it as one of their most important strategic priorities. In this study, we focus on the increasing global competition in semiconductor manufacturing by introducing a mathematical model wherein each country faces a two-stage stochastic optimization model to enhance its semiconductor manufacturing capabilities. In the first stage, countries prepare for a potential global event by boosting production capacities; in the second stage, they manage procurement strategies considering the event's impact on global supply chains and resource availability. Using the concept of variational inequalities, we develop an algorithm to solve several numerical examples that include various scenarios and strategies. Through our analysis of these examples, we provide insights into how different factors and decisions influence the outcomes of global competition in semiconductor manufacturing.

TB35

Summit - 427

TSL Mid-Career and Lifetime Achievement Awards

Invited Session

Transportation Science and Logistics (TSL)

Chair: Hani Mahmassani, Northwestern University, Evanston, IL, 60208, United States

Co-Chair: Karen Smilowitz, Northwestern University, Evanston, IL, United States

1 - TSL Mid-Career Award Winner Talk

Alejandro Toriello, Georgia Tech ISyE, Atlanta, GA, United States

I will discuss my work over the last decade on operations and planning for same-day delivery, as well as a few other highlights.

2 - Reflections

Martin Savelsbergh, Amazon Australia, Darlington, Australia

I will reflect on my 40 years of research in the area of vehicle routing and scheduling.

TB36

Summit - 428

Network Design and Optimization

Invited Session

TSL: Facility Logistics

Chair: Mojtaba Hosseini, University of Iowa, Iowa City, IA, United States

1 - Exact Solution Method for Multi-Stakeholder Freight Transportation Systems under Uncertain Demand

Mojtaba Hosseini, University of Iowa, Iowa City, IA, United States, Gita Taherkhani, Ali Hassanzadeh

This study focuses on investigating the tactical planning of an integrated multi-stakeholder system. The system receives time-dependent requests from carriers and shippers and optimizes in time and space the operations and transportation activities through the consolidation of loads of different shippers into the same vehicles. The demand values of shippers are taken under uncertainty. The aim of tactical planning in this system is to build an efficient service network and schedule. to satisfy the demand and requirements of shippers by making use of the predicted services and their capacities offered by the carriers. A two-stage stochastic program is presented and an exact decomposition-based algorithm is developed. Extensive computational analysis is performed to evaluate the impact of uncertainty on the solution of the proposed model.

2 - Effects of route-related safety risk factors on the optimal location of delivery hubs for an urban logistics network

Jihun Kang, Inha University, Incheon, Korea, Republic of, Junsu Kim, Jonghoo Lee, Hosang Jung

The acceleration of urbanization and the continuous development of e-commerce have concentrated population and logistics demands in urban areas, thereby increasing the complexity of urban logistics networks. This rise in complexity has heightened safety risks, exacerbated by logistics facilities and freight trucks, which negatively affect nearby residents and commuting traffic. In response, this study proposes a p-median based optimization model to address the delivery hub location problem, incorporating economic considerations and six safety risk factors for each route. These six factors, identified through a review of relevant literature, are traffic congestion, road accessibility, safety zones, driver's visibility, left turns at intersections, and density of junctions. The objective is to select a delivery hub that minimizes the weighted sum of distances and safety risk scores, calculated using these factors. Using real data from the Korea postal network and demand locations (post offices) in Seoul, we assign travel distances and safety risk scores between candidate delivery hubs in the Seoul Metropolitan Area. A sensitivity analysis is conducted to explore how varying the weights of economic costs and safety risk scores affects the optimal location. Our findings underscore the significant impact of safety risk factors on delivery hub selection and delineate optimal trade-offs between economic and safety considerations.

3 - Sensor-Driven Predictive Vehicle Maintenance and Routing Problem with Time Windows

Iman Kazemian, Wayne state university, Detroit, MI, United States, Murat Yildirim, Bahar Cavdar

Advancements in sensor technology have revolutionized how we manage the maintenance of vehicle fleets by providing better predictions of vehicle-specific failures. This change shows a big shift in how we predict failures, highlighting the important role of sensor data in addressing risks during operations. However, it is often challenging to transition from vehicle-specific predictions to practical applications in operations and routing. This paper addresses this problem by formulating an optimization problem in the presence of complex operational interactions. A new framework has been developed, integrating sensor-driven vehicle prognostics with a single-vehicle routing problem with time windows. To quantify the impact of failure risks predicted by sensor data, we use an adaptive maintenance cost function that changes dynamically as a function of remaining useful life (RUL) distributions. This approach helps to balance early maintenance, which can waste the remaining useful life, and delayed maintenance, which increases the risk of failure. Routing problems with time windows are inherently challenging, and integrating maintenance considerations adds significantly to its computational complexity. To address this, we have developed the 'Iterative Alignment Method' a new heuristic that builds upon the iterative solution techniques of the Lin-Kernighan-Helsgaun (LKH) algorithm. This approach generates high-quality solutions and enables the handling of large-scale problems efficiently. Moreover, compared to the traditional periodic maintenance strategy, our method shows improvements in operational and maintenance costs as well as in overall fleet reliability.

TB37

Summit - 429

Advancements in Ride-Sharing System Optimization

Contributed Session

Chair: Shenyang Chen

1 - Non-myopic Dispatch in On-Demand Ride-Pooling Systems Using Reinforcement Learning

Farnoosh Namdarpour, New York University, Brooklyn, NY, United States, Joseph Chow

Ride-pooling, also known as ridesharing, shared ride-hailing, or microtransit, is a service wherein passengers share rides. Passengers are typically assigned to vehicles through a centralized fleet dispatch algorithm, allowing for pickups and drop-offs along the route. This service can reduce costs for both passengers and operators and reduce congestion and environmental impacts. However, a common limitation of such services is their myopic decision-making approach, ignoring the long-term consequences of dispatch decisions made at any given point. One promising solution to address this challenge is the application of reinforcement learning (RL). While RL has been extensively studied in the context of ride-hailing systems, there is a scarcity of literature exploring its use in ride-pooling systems. In this study, we build upon the learning and planning framework proposed by Xu et al. for ride-hailing systems, extending it to enable non-myopic decision-making in ride-pooling systems. By employing n-step temporal difference learning, we derive spatiotemporal state values, and subsequently evaluate the effectiveness of the non-myopic policy using NYC taxi data. Results demonstrate that the non-myopic policy leads to significantly lower rejection rates and shorter travel time for passengers at the cost of slightly longer traveled distances by vehicles compared to the myopic policy. Furthermore, in addition to enhancing performance metrics, the implementation of the non-myopic policy can substantially decrease fleet size, thereby offering significant cost savings for operators.

2 - Real-Time Large-Scale Ridesharing with Flexible Meeting Points

Zuhayer Mahtab, University of Southern California, Los Angeles, CA, United States, Maged Dessouky

Rideshare systems can increase the efficiency of the transportation system in large metropolitan areas of the United States. It increases efficiency by providing flexible and convenient commutes to passengers, thereby reducing the number of solo drivers. For a real-time rideshare system to be effective, drivers and passengers need to be matched quickly with minimal waiting time. Also, drivers need to be provided with routes that will minimize their detours and the total traveling cost of the system. This matching and routing system needs to be scalable to implement a rideshare system on a city-wide scale.

Dynamic ridesharing system performance can be improved by forecasting future demands from historical data. Also, the rideshare routes provided to drivers may need to be updated in real time as traffic conditions change. In this project, we develop a framework that, in real-time, matches drivers to passengers and route rideshare drivers that incorporates traffic data and future demands to provide improved routes. Our solution framework uses traffic simulation software to achieve this goal. We also incorporate common pickup and drop-off points for multiple passengers, potentially reducing travel costs and time. Additionally, we include an incentive system that will reduce passengers' traveling costs, promoting the adoption of rideshare systems. Our proposed framework is scalable to large metropolitan areas' demands.

3 - Achieving Balanced Pickup Waiting and Detour Delay: An Analytical Model for Ride-Sharing Based on Dynamic Bipartite Matching

Shenyang Chen, University of California-Davis, Davis, CA, United States, Michael Zhang

Spontaneous ride-sharing services have more room for improvement as they compensate for the insufficient supply of e-hailing taxis. The origin-matching strategy, commonly used by e-hailing taxis to minimize pickup waiting time, can lead to considerable detour delay if directly applied to ride-sharing services. Improving the service quality of ride-sharing systems requires balancing pickup waiting time, detour delay, and vehicle mileage savings. This study developed an analytical model to investigate the effectiveness of an ellipsoid matching strategy in balancing riders' pickup waiting time and detour delay. Based on the dynamic bipartite matching algorithm, this strategy incorporates riders' origins and destinations into the matching cost calculation, thereby reducing detour distances.

Regarding the competition among idle vehicles and pool-seeking vehicles, we compare the performance measures of simultaneous matching and sequential matching (i.e., riders are matched to pool-seeking vehicles first and then to idle vehicles). By deriving closed-form formulas of matching probability and expected travel distance, we provide a reasonable estimation to evaluate the impact of demand and supply on the rider's average service time, including pickup waiting time and onboard travel time. The numerical experiments demonstrate that the ellipsoid matching strategies offer practical benefits, showing a significant reduction in detour delay and a minor increase in pickup waiting compared to the origin-matching strategy. Specifically for ellipsoid matching, we find that simultaneous matching outperforms sequential matching in terms of vehicle mileage savings.

TB38

Summit - 430

Surface and Airport Operations

Contributed Session

Aviation Applications

Chair: Mohammad Reihaneh, IESEG School of Management, La Garenne Colombes, France

1 - Dynamic Departure Pushback Control at Airports: Hybrid Models, Simulations, and Policies

Jita Desai, IIM-Bangalore, Bangalore, India, Sandeep Srivathsan, Guan Lian

Airport surface congestion can lead to significantly long taxi-out times, increased fuel-burn costs, and excessive emissions of greenhouse gases. To curtail these undesirable effects, we propose multiple hybrid penalty-based dynamic pushback control (HPDPC) policies that employ penalty functions dependent on both the taxiway queue limit and the current queue length to ration the pushback frequency at airports, and trades taxiway queueing times with gate-hold delays to minimize the total operational cost (fuel-burn and gate-hold costs). Using data from Beijing Capital International (PEK) airport, the performance of the proposed policies are compared with that of the no-control (baseline) policy, the traditional K-control policy, and the recently proposed penalty-based dynamic pushback control policy. Detailed Monte Carlo simulations and analytical models that study the sensitivity of the total cost function to various problem parameters are presented, and our results indicate that deploying a HPDPC policy that is more stringent than the K-control policy but more flexible than the PDPC policy can be beneficial. Additionally, we also show that deploying a very stringent policy can be detrimental to overall system performance, thereby validating the practical efficacy of the proposed HPDPC policies.

2 - Optimizing Aircraft Taxi Processes: A Multi-Server Queueing Model with Cost-Minimizing Queue Discipline

Kevin James, Brigham Young University, Provo, UT, United States, David Grimsman

Optimization of the aircraft taxi process can significantly decrease operating costs for air carriers, especially at major airports where taxi processes are multi-phased. To improve the efficiency of this process, we model the problem as a queueing system with numerous servers, to allow for cases where multiple runways act as servers in parallel. Additionally, we propose a novel queue discipline which minimizes the mean cost of delays over all flights at a given airport. As an alternative to FIFO, the proposed discipline accounts for aircraft delays at the gate before the aircraft enters the queueing system. We give a sensitivity analysis of the optimized output of our model to perturbations in the variance of service times and arrival rates. Finally, we demonstrate our results via simulation of an example system operating under normal conditions.

3 - Intelligent Routing of Baggage Handling System via Shortest Path-Guided Reinforcement Learning

Keyu TIAN, City University of Hong Kong, Hong Kong, Hong Kong, Li Zeng

Real-time routing of baggage handling system (BHS) becomes challenging with the rising of air travel demands and the advancements of conveyor systems. Traditional methods for solving the shortest path (SP) problem have been used for routing planning. Recently, reinforcement learning (RL) was introduced to this field due to its intelligence in dynamic decision making. However, the vast sample space required for the RL model and the lengthy training time resulted pose an obstacle to implementing this method in practice. This study proposes a novel method called the shortest path-guided reinforcement learning (SP-guided RL) to solve this problem. We adopt a double deep Q-network for RL and use shortest path strategy to direct the training of RL to reduce the exploration sample space and save training time. Furthermore, to enhance the robustness of the model to variations in baggage handling practice, we expand the conventional framework of routing planning to include two critical aspects, including the optimal arrangement of airlines in the BHS and data generation based on limited observations using the Wasserstein conditional generative adversarial network (WCGAN). Simulation study and case study validate the superior performance of the proposed method over existing methods and provide insights about the effect of SP guidance in optimal routing.

4 - Improving Efficiency of Aircraft Turnaround Time in Airport Ground Operation

Jamorn Wacharasing, Chulalongkorn University, Bangkok, Thailand, Natchaphol Prakobphol, Ponlapat Phataraphruk, Pitaakphong Rattanagraikanakorn

This research addresses the need for efficient Turnaround Time (TAT) management in aviation by modifying the ground handling process, providing actionable insights for enhancing the competitiveness and sustainability of the aviation sector. The systematic approaches involve data collection, solutions design, empirical modeling for performance analysis, and optimization analysis. The significance has been made in reducing TAT, particularly in unloading and loading operations which are highlighted as the most-time consuming tasks. The data collection is made with the observation of 39 samples in two airports including Suvarnabhumi Airport (Bangkok, Thailand) and Changi Airport (Singapore). The proposed solutions offer practical strategies for minimizing operational time of both processes. The empirical models are developed to be used in conducting the performance analysis to ensure the ability of each solution in reducing the operational time. Apart from the performance, the financial and environmental impacts are also considered in optimization analysis. The research found that the optimal solutions are 1. Additional Prepared Dollies for Unloading Process and 2. Containers and Pallets Ordering for Loading Process together can minimize the TAT by 5 minutes per flight which considered as the reduction of carbon emission by 222 tons per year, and it also provide cost saving of 15.79 million US dollars per year with the additional revenue generation of 33.01 million US dollars per year for the airlines by scheduling 565 additional flights. However, challenges which are operational variations and sample size limitation warrant further research.

5 - A branch-cut-and-price algorithm for the aircraft sequencing problem with aircraft holding options

Mohammad Reihaneh, IESEG School of Management, Paris, France, Hamed Reihaneh

Efficient utilization of runways, as major bottlenecks at airports, not only minimizes flight delays but also plays a vital role in reducing fuel emissions and mitigating environmental impacts. Aircraft sequencing problem (ASP) studied in this paper seeks to improve this efficiency through optimally sequencing departure/arrival aircraft on multiple runways. Most of the literature on ASP assumes that arriving aircraft can land at any time in a given time windows through speed adjustment. However, in this paper we study a more realistic variant of ASP where arrival flights can adjust their landing time through either "speed adjustment" or "airborne holding". This ASP problem is formulated as a set-partitioning model and solved using a branch-cut-and-price (BCP) algorithm. The proposed algorithm is also compared against the state-of-the-art algorithm for the base ASP (without holding patterns). Computational results show that the new algorithm is in several orders of magnitude faster and capable of solving larger instances within the given time limit.

TB39

Summit - 431

Green House Gas Reduction in the North American Railway Industry

Invited Session

Railway Applications

Chair: Xuesong (Simon) Zhou, ASU, 11235 E Beryl Ave, Tempe, AZ, 85259, United States

Co-Chair: Tyler Dick, University of Texas at Austin, Austin, TX, United States

1 - Quantitative Evaluation and Optimization of the Potential Benefits of Battery Electric Locomotives with ALTRIOS

Diwen Shi, University of Texas at Austin, Austin, TX, United States

In an effort to decarbonize, North American freight railroads are exploring alternative energy technologies such as the Battery Electric Locomotive (BEL). A BEL produces tractive effort from electricity stored in onboard batteries that are re-charged at terminal stations and/or through regenerative braking. Because of their dependance on braking energy, the performance, range, and diesel fuel savings of BELs are expected to be highly variable between particular routes, train types and operating plans. To investigate these potential sensitivities, the diesel fuel savings obtained by adding BELs to conventional diesel-electric locomotive consists is simulated with the Advanced Locomotive Technology and Rail Infrastructure Optimization System (ALTRIOS). Using validated mathematical formulas developed by the research team to describe freight train, conventional locomotive, and BEL performance, ALTRIOS simulations provide detailed energy profiles across a range of BEL deployment scenarios. Analysis of terminal charging strategies and charger power rates reveals a trade-off between battery size, charge rate and annual energy savings. While general trends are observed, BEL implementations must carefully match battery capacity and charging strategy to operational needs on a case-by-case basis, instead of standardizing across all routes. This observation leads to a challenging optimization problem of efficiently deploying different size BELs to specific train runs and topography. Overall, BELs show promising model results in terms of reducing the diesel fuel consumption and emissions of mainline freight railroad operations. The various mainline deployment strategies investigated here demonstrate the potential versatility in BEL application by freight railroad practitioners outside of limited yard and terminal service.

2 - IMAGINED: Inter Modal Analytics for Green Infrastructure Network Energy Decarbonization

Hani Mahmassani, Northwestern University, Evanston, IL, United States, Pablo Durango-Cohen, Adrian Hernandez, Craig Philip, Hiba Baroud

We describe the development of an online, open-source platform to support the (re)design and operation of the nation’s multimodal freight transportation and logistics system. The objectives are to inform/optimize the transition from existing fossil fuels to alternative/renewable energy sources, and to improve the resilience freight transportation systems to disruptions. The platform consists of interrelated modules to support evaluation, selection, and roll out of infrastructure investments across the nation’s multimodal freight network encompassing roadways, railways, and waterways, including the facilities and vehicles that connect them and ensure their safe and efficient operations. Because transloading terminals play crucial roles in determining the feasibility and efficiency of intermodal operations, and are thus essential nodes in decarbonizing the entire system, the platform includes a microsimulation module focusing on these hubs. In addition, the platform includes modules of interest to shippers and carriers to support freight routing on intermodal corridors, and to support the renewal of vehicle fleets.

3 - Intermodal routing on road and rail with comparisons of green-house gas emissions

Marc Meketon, Oliver Wyman, Princeton, NJ, United States

We will describe on-going work of efficient routing of intermodal traffic - using both truck and rail (and possibly water routes within the U.S.), and the calculations of resulting green-house-gas emissions.

4 - Energy and economic impacts of advanced freight transportation concepts

Olcay Sahin, Argonne National Laboratory, Lemont, IL, United States

In 2021, the transportation sector significantly contributed to greenhouse gas emissions. Medium and heavy-duty vehicles (MHDVs), primarily responsible for U.S. freight movement and expenses, pose environmental challenges. In line with the U.S. goal of net-zero carbon emissions by 2050, advanced freight transportation concepts play an important role energy and economic impacts. Adoption of alternative fuels such as hydrogen, biofuels, or electricity can reduce reliance on fossil fuels, thereby decreasing greenhouse gas emissions and mitigating environmental impact. Improved efficiency in freight transportation often leads to lower operational costs for businesses, including reduced fuel expenses and maintenance costs. This can result in higher profit margins or more competitive pricing for consumers. To quantify the potential of advanced freight transportation concepts when deployed at scale at the system level, the DOE SMART Mobility workflow is implemented (e.g., POLARIS, SVTrip, Autonomie), leveraging multiple metrics including mobility, energy, emissions, and cost.

TB40

Summit - 432

Behavioral Perspective of Operations: Adapting to a Changing World

Invited Session

Behavioral Operations Management

Chair: Lei Hua, University of Texas at Tyler, Tyler, TX, United States

1 - The Market Value of Social Responsibility Communications: a Cross-Cultural Behavioral Study

Lei Hua, University of Texas at Tyler, Tyler, TX, United States, Tim Kraft, Doug Thomas, Yanchong Zheng

With consumers' increasing expectation for transparency, companies need to understand how to effectively communicate information about their supply chain practices. Our study examines consumers' valuations of sustainability communication and the impact of different social responsibility topics and message formats on these valuations.

2 - Assessing Consumer Reactions to Sustainable Supply Chain Disclosures: A Comparative Study Using Likert Scales and Brain Computer Interfaces

Yanji Duan, University of North Florida, Jacksonville, FL, United States, Qingyun Zhu, Feng Liu

In the evolving landscape of sustainable operations, businesses are increasingly adopting practices that prioritize environmental and social responsibility. Central to this movement is the concept of sustainability supply chain disclosure, which enhances transparency and accountability. By openly communicating their sustainability initiatives along their supply chains, firms enable consumers, who are increasingly driven by sustainability values, to make informed purchasing decisions. Within this context, cognitive appraisal theory offers a framework to understand consumer reactions to sustainability disclosures. It suggests that consumers' cognitive evaluations of products and services influence their emotional and behavioral responses, affecting decisions and behaviors like purchase intentions and brand attitudes. However, the primary method for assessing these emotions, the Likert scale, has limitations due to its reliance on self-reporting and the subjective nature of emotions. This research proposes to bridge this gap by combining traditional Likert scales with Electroencephalography (EEG) to capture genuine, real-time emotional responses. This innovative approach seeks to answer key questions about the consistency of results between Likert scales and EEG measurements, the precision of EEG in revealing emotional responses, and ways to enhance Likert scales using EEG findings. This study will conduct two behavioral experiments, one online scenario-based and the other in a lab using EEG, to examine consumer responses to sustainable supply chain disclosures.

3 - A Data-Driven Approach to Assortment Planning for Collective-Intelligence Crowdsourcing Contests with Heterogenous Participant Motivations

Olumurejiwa Fatunde, University of Toronto, Toronto, ON, Canada, Gonzalo Romero

Many organizations have increasingly adopted novel methods of organizing distributed workers, including crowdsourcing. This paper studies how strategically-selected contest offerings that incorporate user motivation can improve crowdsourcing for skilled work in a setting focused on collective, rather than individual, performance. We explore two questions: first, we seek to learn from historical data the underlying motivations that drive participation on distributed-task platforms. Second, we determine the optimal assortment of contests to offer each motivation-based class. We use data from a healthcare platform that runs rank-order tournaments to collect and aggregate diagnostic opinions, and that can tailor contest offerings and prizes to attract users and shape behavior. As medical misdiagnosis can have fatal consequences, innovative models that improve diagnostic accuracy or facilitate incorporation of external expertise can transform healthcare. We borrow from the dynamic assortment planning literature by framing the contest offering decision as an assortment problem with independent search, perishable goods, and dependent demand. We treat motivation as a latent variable and apply a finite mixture model to historical performance data to understand whether users are primarily motivated by learning opportunities, monetary prizes, or affinity for competition. We determine the optimal assortment, balancing motivation-specific user utility and platform goals. Data-driven estimates provide greater insight into user motivation than self-reported responses and reduce the number of contests offered by 38%, simplifying the user experience without increasing cost. Users who had previously entered contests included in their streamlined assortments achieved above-average performance, suggesting the ability to systematically eliminate poor performance.

4 - Fair Funnels: How Hiring Processes (and Interventions) Impact Di-Versity and Quality

Eunjee Kim, UW-Madison, Madison, WI, United States, Jordan Tong, Xiaoyang Long

Behavioral research has uncovered several biases that evaluators display when judging quality. How do these biases impact the diversity and quality of hires? We examine this question using a stylized model of recruiting structures, i.e., the funnel shape from the candidate pool to the final hire(s), and the hiring committee composition. We show that the impact of judgment biases depends critically on the recruiting structure. Whereas a basic model of taste-based discrimination leads to intuitive dynamics, incorporating other established biases like outgroup homogeneity and judgment noise can lead to surprising results. We deliver insights into what combinations of judgment biases and recruiting structures most (dis)advantage minorities, and how organizations can change their recruiting structure to improve both quality and diversity.

TB41

Summit - 433

Incentives and Information Design in Markets

Invited Session

Applied Probability Society

Chair: Prakirt Jhunjhunwala, Columbia Business School, 601 West 113th st , New York, NY, 10025, United States

Co-Chair: Yash Kanori, Columbia, New York City, NY, United States

1 - Markov Persuasion Processes with Endogenous Agent Beliefs

Krishnamurthy Iyer, University of Minnesota, Minneapolis, MN, United States, Haifeng Xu, You Zu

We study a Markov persuasion process, where a long-lived principal ("sender") persuades a stream of short-lived agents by sharing information about an evolving state. The state transitions are Markovian and the sender seeks to maximize the long-run average reward by committing to a (possibly history-dependent) signaling mechanism. While most previous studies of Markov persuasion consider agents with exogenously specified beliefs, we study a more natural variant with endogenous beliefs that depend on the chain's realized history. A key challenge in studying this variant is to model the agents' partial knowledge about the history. We focus on settings where each agent observes the history with a k period lag. We show that the sender's long-run average payoff (weakly) increases with the lag, and provide efficient LP formulations to compute the optimal signaling mechanisms for k = 0 (full history information) and k = infinity (no history information). We also find sufficient conditions under which sender's optimal payoff is the same regardless of the lag with which the agents observe the history. For general values of the lag, we identify challenges in formulating the sender's problem, and provide a bilinear optimization framework to solve a restricted problem. Finally, using robust persuasion, we design simple signaling mechanisms that are approximately optimal when the lag is large.

2 - Incentives for Exploration at Market Equilibrium

Vijay Kamble, University of Illinois Chicago, Chicago, IL, United States

Online marketplaces face an exploration problem: the qualities of new supply units are unknown and must be discovered through customer feedback so that higher-quality supply gets prioritized for matching. However, customers are generally myopic and unwilling to participate in exploratory matches, leading to the well-known concern of incentivizing such exploration. This paper uncovers the role of competitive pricing effects arising from market congestion in creating incentives for exploratory behavior among myopic customers. The intuition is that since established higher-quality supply units are expected to be more popular and, hence, more congested, they naturally demand higher prices at a competitive equilibrium than new supply units, effectively incentivizing customers to explore. This paper presents a comprehensive analysis of the extent to which this intuition holds and exogenous incentives for exploration are necessary for such markets. To investigate this question, we define a competitive equilibrium notion for markets with evolving public information about supply units. The key result establishes that, under a tightly characterized market regularity condition, the ratio of equilibrium matching value to the system optimal value is bounded by the aggregate congestion level, indicating that congested regular markets inherently incentivize exploration. Furthermore, we show that in markets with linear information transition structures, the equilibrium achieves the first-best system-optimal value regardless of the congestion level. Finally, we address the problem of designing optimal price interventions to align market equilibrium with the system-optimal solution. Overall, our results inform market designers grappling with the concern of incentivizing exploration in various online marketplaces.

3 - Information Design for Spatial Resource Allocation

Manxi Wu, Cornell University, ORIE, ITHACA, NY, United States, Ozan Candogan

In this paper, we study platforms where resources and jobs are spatially distributed, and resources have the flexibility to strategically move to different locations for better payoffs. The price of the service at each location depends on the number of resources present and the market size, which is modeled as a random state. Our focus is on how the platform can utilize information about the underlying state to influence resource repositioning decisions and ultimately increase commission revenues. We establish that in many practically relevant settings a simple monotone partitional information disclosure policy is optimal. This policy reveals state realizations below a threshold and above a second (higher) threshold, and pools all states in between and maps them to a unique signal realization. We also provide algorithmic approaches for obtaining (near-)optimal information structures that are monotone partitional in general settings.

4 - Optimal Signaling Mechanism under Demand Variation and Customers with Heterogeneous Patience

Prakirt Jhunjhunwala, Columbia Business School, New york, NY, United States, Jing Dong, Yash Kanoria

In an unobservable queue, where customers lack complete wait time information, a throughput-maximizing server aims to exploit the information asymmetry by strategically signaling coarse congestion information to incentivize customers' arrival into the system. In this work, we characterize the structure of the optimal information-sharing mechanism used by the server by adopting a Bayesian persuasion framework to model the customers' response to the provided signal. We incorporate two important real-world features, introducing additional information asymmetries between servers and customers. Firstly, we explore customer heterogeneity with private types, where customers vary in willingness to wait, modeled using the utility function. Secondly, we consider demand variation, where the customer arrival rate varies from one day to the next and is only observable to the server. We model the server's throughput optimization problem as an infinite dimensional linear program in terms of steady-state distribution. We also develop an equivalent Constrained Markov Decision Process (C-MDP) formulation. Both formulations offer insights into the properties of the signaling structure.

Customer heterogeneity works in favor of the customers, allowing some customer classes to receive strictly positive utility under private customer types. We discover a counter-intuitive phenomenon where the optimal signaling mechanism has a laminar structure; for example, with two customer types, the interval of queue lengths in which only patient customers enter is nested within the queue length intervals in which all customers enter. In contrast to customer heterogeneity, demand variation favors the server by providing additional flexibility due to the customer's lack of accurate arrival rate information.

TB42

Summit - 434

Rare Events and Large Deviations

Invited Session

Applied Probability Society

Chair: Fiona Sloothaak, Eindhoven University of Technology, Eindhoven, Netherlands

1 - From Heavy-Tailed Rare Events to Global Dynamics

Chang-Han Rhee, Northwestern University, Evanston, IL, United States, Xingyu Wang

We discuss how heavy-tailed large deviations asymptotics can characterize the nominal global dynamics.

2 - On the first passage times of spatial branching processes in R^d

Zhenyuan Zhang, Stanford University, Stanford, CA, United States, Jose Blanchet, Wei Cai, Shaswat Mohanty

Spatial branching processes, including branching Brownian motion, branching random walks, and their variants, have a long history in probability and a wide range of applications in ecology, population biology, and modeling epidemics. In this talk, we initiate the study of their first passage times (FPT), by considering the FPT to a ball of radius one with a distance x from the origin, in a Euclidean space of an arbitrary dimension. We establish precise asymptotics of the first passage times as a function of x. We also discuss interesting consequences in polymer physics. This is based on joint work with Jose Blanchet, Wei Cai, and Shaswat Mohanty.

3 - Decision-Making under Extreme Risks: Configuring Optimization Algorithms for Rare-Event Optimization

Wasin Meesena, Columbia University, New York, NY, United States, Henry Lam

We consider stochastic optimization where the goal is not only to optimize an average-case objective, but also mitigate the occurrence of rare but catastrophic events. This problem, which is motivated from emerging applications such as safe AI, requires an integration of variance reduction into sampling-based optimization algorithms. These variance reduction techniques, widely studied for estimation tasks, helps sharpen the accuracy of the associated empirical estimates in the algorithm with the hope of guaranteeing a realistic optimization runtime. Despite the reasonableness of this approach, we demonstrate how a variance-reduction-optimization integration, even executed in an adaptive fashion studied by recent works, encounters fundamental challenges when using stochastic gradient descent procedures. On a high level, the challenge arises from the extreme sensitivity of tail-based objectives with respect to the decision variables, which renders the failure of traditional Lipschitz-based analyses. We offer some potential remedies and supporting numerical results.

4 - Scale-free cascading failures: Generalized approach for all simple, connected graphs

Agnieszka Janicka, Technical University Eindhoven, Eindhoven, Netherlands, Fiona Sloothaak, Maria Vlasiou

Cascading failures, wherein the failure of one component triggers subsequent failures in complex interconnected systems, pose a significant risk of disruptions and emerge across various domains. Understanding and mitigating the risk of such failures is crucial to minimize their impact and ensure the resilience of these systems. In multiple applications, the failure processes exhibit scale-free behavior in terms of their total failure sizes. Various models have been developed to explain the origin of this scale-free behavior. A recent study proposed a novel hypothesis, suggesting that scale-free failure sizes might be inherited from scale-free input characteristics in power networks. However, the scope of this study excluded certain network topologies. Here, motivated by power networks, we strengthen this hypothesis by generalizing to a broader range of graph topologies where this behavior is manifested. Our approach yields a universal theorem applicable to all simple, connected graphs, revealing that when a cascade leads to network disconnections, the total failure size exhibits a scale-free tail inherited from the input characteristics. We do so by characterizing cascade sequences of failures in the asymptotic regime.

TB43

Summit - 435

Microeconomics, from a Computer Science Perspective

Invited Session

Applied Probability Society

Chair: Kangning Wang, 94305

1 - Deterministic Budget-feasible Clock Auctions

Daniel Schoepflin, Rutgers University, New Brunswick, NJ, United States

We revisit the well-studied problem of budget-feasible procurement, where a buyer with a strict budget constraint seeks to acquire services from a group of strategic providers (the sellers). During the last decade, several strategyproof budget-feasible procurement auctions have been proposed, aiming to maximize the value of the buyer, while eliciting each seller's true cost for providing their service. These solutions predominantly take the form of randomized sealed-bid auctions: they ask the sellers to report their private costs and then use randomization to determine which subset of services will be procured and how much each of the chosen providers will be paid, ensuring that the total payment does not exceed the buyer's budget. Our main result in this paper is a novel method for designing budget-feasible auctions, leading to solutions that outperform the previous auctions in multiple ways. First, our solutions take the form of clock auctions, and thus satisfy a list of very appealing properties, such as obvious strategyproofness, group-strategyproofness, transparency, and unconditional winner privacy, making these auctions much more likely to be used in practice. Second, in contrast to previous results that heavily depend on randomization, our auctions are deterministic. We provide an affirmative answer to one of the main open questions in this literature, asking whether a deterministic strategyproof auction can achieve a constant approximation when the valuation function is submodular over the services. This is joint work with Eric Balkanski, Pranav Garimidi, Vasilis Gkatzelis, and Xizhi Tan and originally appeared in a paper accepted at SODA 2022.

2 - Price Discrimination under Price Regulations

Yiheng Shen, Duke University, Durham, NC, United States, Kamesh Munagala, Renzhe Xu

In this presentation, I will talk about our recent study on third-degree price discrimination in a model first presented by Bergemann, Brooks, and Morris [2015]. Since such price discrimination might create market segments with vastly different posted prices, we consider regulating these prices, specifically, via restricting them to lie within an interval. Given a price interval, we consider segmentations of the market where a seller, who is oblivious to the existence of such regulation, still posts prices within the price interval. We show the following surprising result: For any market and price interval where such segmentation is feasible, there is always a different segmentation that optimally transfers all excess surplus to the consumers. In addition, we characterize the entire space of buyer and seller surplus that are achievable by such segmentation, including maximizing seller surplus, and simultaneously minimizing buyer and seller surplus.

This is joint work with Kamesh Munagala and Renzhe Xu.

3 - Breaking the Metric Voting Distortion Barrier

Prasanna Ramakrishnan, Stanford University, Stanford, CA, United States, Moses Charikar, Kangning Wang, Hongxun Wu

We consider the following well-studied problem of metric distortion in social choice. Suppose we have an election with voters and candidates who lie in a shared metric space. We would like to design a voting rule that chooses a candidate whose average distance to the voters is small. However, instead of having direct access to the distances in the metric space, each voter gives us a ranked list of the candidates in order of distance. Can we design a voting rule that, regardless of the election instance and underlying metric space, chooses a candidate whose cost differs from the true optimum by only a small factor (known as the distortion)? A long line of work culminated in finding deterministic voting rules with metric distortion 3, which is the best possible for deterministic rules and many other classes of voting rules. However, without any restrictions, there is still a significant gap in our understanding. Even though the best lower bound is substantially lower at 2.112, finding a randomized rule that guarantees distortion $3 - \epsilon$ for some constant ϵ has been a major challenge in computational social choice. In this work, we give a rule that guarantees distortion less than 2.753 by carefully randomizing between Maximal Lotteries, a natural game-theoretical voting rule dating back to the 60's, and new voting rules that we introduce. Based on joint work with Moses Charikar, Kangning Wang, and Hongxun Wu.

4 - The Secretary Problem with Predicted Additive Gap

Sherry Sarkar, Carnegie Mellon University, Pittsburgh, PA, United States, Alex Braun

The secretary problem is one of the fundamental problems in online decision making; a tight competitive ratio for this problem of 1/e has been known since the 1960s. Much more recently, the study of algorithms with predictions was introduced: The algorithm is equipped with a (possibly erroneous) additional piece of information upfront which can be used to improve the algorithm's performance. Complementing previous work on secretary problems with prior knowledge, we tackle the following question: What is the weakest piece of information that allows us to break the barrier? To this end, we introduce the secretary problem with predicted additive gap. As in the classical problem, weights are fixed by an adversary and elements appear in random order. In contrast to previous variants of predictions, our algorithm only has access to a much weaker piece of information: an additive gap. This gap is the difference between the highest and kth highest weight in the sequence. Unlike previous pieces of advice, knowing an exact additive gap (provably) does not make the problem trivial. Our contribution is twofold. First, we show that for any index k and any gap, we can obtain a competitive ratio of 0.4 (even if we do not know k), hence beating the prevalent bound for the classical problem by a constant. Second, a slightly modified version of our algorithm allows us to prove standard robustness-consistency properties as well as improved guarantees when knowing a range for the error of the prediction.

TB44

Summit - 436

Service Science Best DEIJ/Cluster Paper

Award Session

Service Science

Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, United States

1 - Innovative Resource Allocation in Public Services: Incorporating Utility Heterogeneity and Income Disparity

Siddharth Singh, University College London, London, United Kingdom, Owen Wu

Efficient and equitable allocation of a divisible resource and its associated costs or savings among consumers is a critical issue for many public service corporations, especially when consumers have heterogeneous incomes and private levels of resource utility. The challenge arises from the dual dimensions of consumer characteristics and the coupled allocation problems. Common approaches in practice often ignore utility heterogeneity or income disparity, leaving a significant gap between potential and actual outcomes. We develop and analyze a resource allocation method to bridge this gap. Our model incorporates consumer heterogeneity in both income levels and private utility from the resource. We formulate the problem of allocating the resource and its associated costs or savings, aiming to maximize aggregate welfare. We propose a method that offers consumers income-dependent menus (IDM) of quantity and cost (or savings) options, and uncover key structural properties of these menus. Our IDM approach significantly outperforms the considered alternatives in numerical studies calibrated using real-world data. In the realm of resource allocation where both income and resource utility levels are heterogeneous, significant welfare gains can be realized by judiciously leveraging the dual dimensions of heterogeneity. Implementing our IDM approach ensures that consumers receive and contribute to resources in a manner that reflects their financial capacities and utility levels, maximizing overall welfare.

2 - Racial and Gender Biases in Customer Satisfaction Surveys: Evidence from a Restaurant Chain

Masoud Kamalahmadi, University of Miami, Coral Gables, FL, United States, Qiuping Yu, Yong-Pin Zhou

Racial and gender inequalities are ubiquitous in the workplace. Whereas previous studies have primarily focused on employer discrimination, we study the role of customers, using 1.5 million transactions and 260,000 customer satisfaction surveys from a full-service casual-dining restaurant chain in the US. We find that customer ratings of servers are biased against racial minority, and, interestingly, also against female servers despite their majority in this occupation. We further find evidences that suggest that statistical discrimination is the primary driver for racial biases, while status-based discrimination is likely the main driver for gender biases. Given these underlying mechanisms, we propose tailored strategies to mitigate the biases.

3 - Online Fair Allocation of Perishable Resources

Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States, Sean Sinclair, Siddhartha Banerjee

We consider a variant of the canonical online fair allocation problem: a decision-maker has a budget of perishable resources to allocate over a fixed number of rounds. Each round sees a random number of arrivals, and the decision-maker must commit to an allocation for these individuals before moving on to the next round. The goal is to construct a sequence of allocations that is envy-free and efficient. Our work makes two important contributions toward this problem: we first derive strong lower bounds on the optimal envy-efficiency trade-off that demonstrate that a decision-maker is fundamentally limited in what she can hope to achieve relative to the no-perishing setting; we then design an algorithm achieving these lower bounds which takes as input a prediction of the perishing order, and a desired bound on envy. Given the remaining budget in each period, the algorithm uses forecasts of future demand and perishing to adaptively choose one of two carefully constructed guardrail quantities. We demonstrate our algorithm's strong numerical performance - and state-of-the-art, perishing-agnostic algorithms' inefficacy - on simulations calibrated to a real-world dataset.

4 - The Dedicated Docket in U.S. Immigration Courts: An Analysis of Fairness and Efficiency Properties

Daniel Freund, MIT, Cambridge, MA, United States, Wentao Weng

The dedicated docket was introduced by the Biden Administration to reform the immigration system. It creates a separate queue for immigration proceedings where judges are supposed to issue decisions for each case within a target timeframe. Its goals are to improve speed, accuracy, and fairness. Though it meets its first goal, legal advocacy groups report that this comes at the expense of the last. Referring to it as a ``Denial of justice'', they find that cases on the dedicated docket routinely fail to access legal representation, and have a much lower asylum grant rate. Against this backdrop, we study the operational implication of the dedicated docket.
We develop a queueing model wherein a policy maker (PM) routes asylees to the regular or the dedicated docket, and sets a delay target for the latter. Constrained by the target, the court allocates its capacity to minimize the average delay. Immigration lawyers schedule their time between dockets to maximize the rate of successful asylum cases. Compared to a single docket, we show that the dedicated-docket system can Pareto-improve the speed and accuracy. However, we also prove that the dedicated docket system satisfies two natural fairness rules only if it is dominated by the single docket.
Our analysis can inform the public discourse. We prove that the lack of fairness is a fundamental design flaw of the new program. Though delay differentiation enables a surprising efficiency gain, this requires unfairness between. Policy makers and legal advocacy groups should be aware of such tradeoffs.

5 - The Power of Simple Menus in Robust Selling Mechanisms

Shixin Wang, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of

We study a robust selling problem where a seller attempts to sell one item to a buyer but is uncertain about the buyer's valuation distribution. Existing literature shows that robust screening provides a stronger theoretical guarantee than robust deterministic pricing, but at the expense of implementation complexity, as it requires a menu of infinite options. Our research aims to find simple mechanisms to hedge against market ambiguity effectively. We develop a general framework for robust selling mechanisms with a finite menu (or randomization across finite prices). We propose a tractable reformulation that addresses various ambiguity sets of the buyer's valuation distribution, including support, mean, and quantile ambiguity sets. We derive optimal selling mechanisms and corresponding performance ratios for different menu sizes, showing that even a modest menu size can deliver benefits similar to those achieved by the optimal robust mechanism with infinite options, establishing a favorable trade-off between theoretical performance and implementation simplicity. Remarkably, a menu size of merely two can significantly enhance the performance ratio compared to deterministic pricing.

TB45

Summit - 437

Critical issues in Pharmaceutical Value Chains

Invited Session

Health Applications Society

Chair: Hui Zhao, The Pennsylvania State University, University Park, University Park, PA, United States

Co-Chair: Ji Won Moon, Pennsylvania State University, University Park, PA, 16802, United States

1 - Impact of Copay Accumulator Ban: Evidence from National Healthcare Utilization Data

Ji Won Moon, Pennsylvania State University, University Park, PA, United States, Hui Zhao, Susan Lu

The recent federal ruling for a nationwide ban on copay accumulator represents the latest development in a decade-long game of cat and mouse in the pharmaceutical industry. The accumulator ban aims to undo the insurer’s restrictions on copay assistance programs designed to promote sales of pricey brand drugs among commercially insured patients. The ban poses uncertainties regarding ensuing market responses that have not been studied. We address the following questions: What impact does the ban have on drug purchase quantity, patient and insurer cost burdens, and the total volume of patients on therapy? What is the ban’s impact on health equity? What stakeholder responses does the ban trigger, and what do they imply about its long-term consequences?

2 - Does Production Location Impact Medication Quality? An Empirical Study on the Effectiveness Variations

Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Xinyu Liang, Ravi Anupindi, In Joon Noh

Global production offers promising avenues for cost reduction but introduces challenges in maintaining consistent quality standards. We study the health effects of medications produced in over 1,000 pharmaceutical manufacturing facilities with over ten million annual prescription patients. Utilizing a large-scale dataset that includes patient-level and production facility-level data, we assess the impact on patients with medication prescriptions for nine conditions. Our empirical analysis reveals significant outcome variations across different manufacturers and production countries and explores potential mechanisms behind these variations. Our study offers insights for regulatory bodies on optimizing inspection procedures and ensuring consistent quality control across borders to safeguard public health.

3 - Pharmaceutical Supply Chain Design Considering Export Bans Induced by Capacity Strains

Martha L. Sabogal De La Pava, Clemson University, Clemson, SC, United States, Emily Tucker

For decades, pharmaceutical supply chains have faced persistent challenges in meeting domestic and global drug demand. Challenges include exposure to multiple types of disruption risks, which are capable of causing failures in the supply chain operation. These failures may lead to drug shortages, endangering patients' health and increasing the costs of healthcare systems. An emerging disruption risk, export bans, has increased due to recent geopolitical instability and persistent capacity strains. We propose a stochastic mixed integer program for global supply chain design that considers disruption risks, such as export bans, correlated natural disasters, and facility quality issues. Global capacity strains induce export bans. A natural disaster may affect countries in proximity. Operational strains such as uncertainty in demand and supplies are also included. We analyze the impact of such disruption risks on optimal supply chain structures and global drug shortages.

4 - Assortment Optimization of Prescription Coupons for Online Discount Programs

Yishen Cai, University of Miami Herbert Business School, Miami, FL, United States, Nan Yang, Hongmin Li

We consider an assortment problem motivated by online prescription-drug discount programs. Coupon vendors such as GoodRx (GDRX) provide discount card for prescription drugs by partnering with Pharmaceutical Benefit Managers (PBMs) and leveraging PBMs' negotiated prices with retailers and manufacturers. GoodRx uses its online platform to list discounted prices for prescription drugs at each participating pharmacy and direct individual patients to retail pharmacies through its listing. We study a coupon vendor's assortment problem of assigning available coupons at each retail pharmacy. The optimal decision must account for patients' choice among different retail pharmacies and balance the coupon vendor's commission revenue with patient benefit. The assortment problem is applicable to other digital marketers such as RetailMeNot, CapitalOneShopping and Goodshop. We propose an innovative and efficient solution approach for the optimal assortment and identify useful insights for decision support.

TB46

Summit - 438

Models for Hospital Operations

Invited Session

Health Applications Society

Chair: A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States

1 - Hospital Versus Home Care: Trading off Pre- and Post-Discharge Infection and Mortality Risks

Mor Armony, New York University, New York, NY, United States, Galit Yom-Tov

Previous research has shown that early discharge of patients may hurt their medical outcomes. However, in many cases the "optimal" length of stay (LOS) and the best location for treatment of the patient are not obvious. A case in point is hematology-oncology (blood cancer) patients, for whom these are critical decisions. Patients with blood cancer are susceptible to life-threatening infections after treatment. Sending these patients home early minimizes infection risk, while keeping them longer for hospital observation minimizes mortality risks if an infection occurs. We develop LOS optimization models for hematology-oncology patients that balance the risks of patient infection and mortality.

We develop a Newsvendor-style model to explore the impact of the infection and mortality risks on the optimal LOS from a single-patient perspective. We further consider the social optimization problem in which capacity constraints limit the ability of hospitals to keep patients for the entirety of their optimal LOS. Using a fluid model, we find that the optimal solution is a priority policy that is based on a speedup index. This policy may block some patients and immediately route them to home care, or speed up some patients' LOS and send them to be home-cared early after an observation period at the hospital.

In a case study, we show that around 75% of the patient population need some observation period. If the hospital is overloaded, using a speedup only policy is optimal for 90% of the patient types; applying it to all patients increases overall mortality risk by 0.5%.

2 - Heterogeneous Effect of Emergency Department Boarding on Inpatient Length of Stay

Xiaole (Alyssa) Liu, New York University Stern School of Business, New York, NY, United States, Jing Dong, Martin Copenhaver, Michael Hu, A. Cecilia Zenteno

Emergency Department (ED) boarding, the practice of holding admitted patients in the ED waiting for an available inpatient bed, is a common issue in many hospitals across the world. It is not only the primary cause of ED crowding but also raises patient safety concerns. When there are limited inpatient beds available, we need to decide how to prioritize the boarding patients in the ED for inpatient admission. In this work, we study the causal impact of boarding on inpatient length of stay (LOS). Using data from around 55,000 ED admissions and leveraging recent developments in double machine learning and causal forest with instrumental variables, we identify patient subpopulations that are affected by boarding differently. Through close collaboration with medical experts, we also identify potential explanations for the heterogeneous effects. This allows us to provide more actionable policy suggestions for boarding management and inpatient bed allocation.

3 - The Burden of Evidence for Operations Research in Hospitals

Martin Copenhaver, Johns Hopkins Hospital and Johns Hopkins University, Baltimore, MD, United States

What evidence is used to make strategic operational decisions in hospitals? In this talk, we will explore this question (and resulting challenges and opportunities) through a variety of examples. In particular, we consider the conceptual divide in medical versus operational decision making, how metrics of success can differ between hospital stakeholders and the academic community, and the implications for the practice of operations research.

4 - The Impact of Provider Behavior on Emergency Department Operations

Huifeng Su, Yale School of Management, New Haven, CT, United States, Lesley Meng, Rohit Sangal, Edieal Pinker

Emergency department (ED) providers play a crucial role in deciding patient admissions or discharges. Such decision-making significantly impacts the utilization of medical resources within the hospital, as it often relies on a range of shared diagnostic resources. Moreover, admitting patients contributes to ED crowding because they occupy ED spaces while awaiting transfer to downstream units. This study explores the behavior of physicians in making diagnostic decisions and the subsequent effects on ED congestion.

5 - Identifying Short-Stay Patients in the Emergency Department

A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States, Thomas Kyle, Martin Copenhaver, Retsef Levi, Peter Dunn, Kyan Safavi, Gohram Baloch, Bethany Daily

Emergency Department overcrowding is a widespread patient safety concern. A significant contributor to this congestion is “boarding” patients who have been admitted but are waiting for an inpatient bed to become available. In this work, we propose establishing a different pathway for ED patients who are predicted to have a short length of stay upon admission (two nights maximum). Our goals are: to minimize disruptions in these patients' care path, to decrease the number of ED boarders, and to open up inpatient beds for more complex patients. We used xgboost to develop a predictive model using patients’ clinical, administrative, and demographic data and validated it with a group of physicians. The model has an out-of-sample AUC of 0.81, a precision of 69%, and is 75% accurate. This algorithm is in the process of being implemented at an academic medical center's ED.

TB47

Summit - 439

Analytics Applications in Hospital Emegency Care

Invited Session

Health Applications Society

Chair: Abdulaziz Ahmed, University of Alabama at Birmingham, Birmingham, AL, 35294-1212, United States

1 - Enhancing Emergency Department Efficiency: Predictive Modeling for Patient Flow Metrics

Abdulaziz Ahmed, University of Alabama at Birmingham, Birmingham, AL, United States, Orhun Vural, Bunyamin Ozaydin, Khalid Aram, James Booth

ED overcrowding poses significant risks, including poor healthcare quality, increased medical errors, higher patient mortality, and low satisfaction. The American College of Emergency Physicians (ACEP) identifies this as a patient flow issue, suggesting that improving flow can mitigate overcrowding. In 2016, ACEP recommended the Full Capacity Protocol (FCP), which involves interventions based on real-time patient flow measures (PFMs) such as the number of patient waiting and boarded patients. This study aims to develop machine learning models to predict different PFMs including total patients in ED waiting rooms, average waiting times, total boarding patients, and average boarding times.

2 - Examining the Impact of Unfinished ED Laboratory Tests on Patient Throughput in Emergency Departments

Lu Wang, Ball State University, Muncie, IN, United States, Yajun Lu, Abdulaziz Ahmed

Laboratory testing during emergency department (ED) visits has been shown to directly influence patients' length of stay (LOS). However, the inappropriate utilization of laboratory resources has become an increasing concern. This study aims to examine the impact of laboratory tests ordered during ED visits but not utilized for physician diagnosis, referred to as "unfinished lab tests by disposition" (ULTD), on ED operations. After addressing potential endogenous factors, we empirically investigate how ULTD affects both ED LOS and inpatient LOS.

3 - Effects of long-stay patients on hospital capacity and flow: An empirical investigation

Esmaeil Bahalkeh, University of New Hampshire, Durham, NH, United States, Tze C Chiam, Yuehwern Yih

Often patients with discharge barriers stay in medical units for non-medical reasons and contribute to flow and capacity issues within and beyond medical units including intensive care units (ICU) and emergency departments. Using patient-level data at two academic hospitals within the same healthcare system, we estimate the impact of long-stay patients on length of stay (LOS) and 30-day readmission of other patients in medical units, as well as boarding time for patients transitioning from both ICU and step-down unit (SDU) to medical unit (MU). Inspired by prior studies on patient flow between ICU and MU, we include occupancy of origin (i.e., ICU, SDU) and destination (i.e., MU) units. We find that a larger proportion of long-stay patients in the MU result in shorter LOS for other patients in the MU, and longer boarding time for patients leaving the ICU to MU. Further, the proportion of long-stay patients is associated with neither 30-day readmission, nor boarding time for patients leaving the SDU. For both the ICU and the SDU, we find that patients experience shorter boarding time leaving these units to the MU as the unit gets busier. Finally, higher occupancy of the MU results in longer boarding time for patients leaving the SDU.

4 - Assessing the Impact of Clinical Symptoms on Emergency Room Capacity During the COVID-19 Pandemic: A Discrete Event Simulation Approach

Osama Aljarrah, Kettering University, Flint, MI, United States, Farnaz Ghazi Nezami

In this study, we have explored the role of clinical symptoms in managing the capacity of emergency room (ER) units during the COVID-19 pandemic by using commercial discrete event simulation (DES) software (FlexSim Healthcare). The focus is on understanding how variations in clinical symptom presentation impact ER operations and patient throughput under pandemic conditions. The ER unit at a hospital is modeled to assess key performance indicators (KPIs) in scenarios with varying degrees of COVID-19 symptom severity and incidence rates. The simulation defines and quantifies operational and capacity management indicators, such as wait times, resource utilization, and patient flow. Compared with a baseline non-pandemic model, multiple scenarios reveal a marked decrease in operational efficiency. Initial findings suggest that the presence of COVID-19 symptoms significantly impacts ER workflows, leading to increased occupancy rates and extended waiting times, thereby straining resources. The study emphasizes the necessity for effective triage protocols and capacity planning strategies tailored to pandemic circumstances to enhance ER responsiveness and mitigate the adverse effects on healthcare delivery. By providing a virtual replication of ER operations during the COVID-19 crisis, this analysis offers valuable insights for healthcare facilities aiming to optimize service levels and patient care in challenging times.

TB48

Summit - 440

Operations Management and Marketing Interface

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: EunSu Lee, New Jersey City University, Jersey City, NJ, 07311, United States

Co-Chair: Hosun Rhim, Korea University Business School, Seoul, N/A

1 - A Conjoint Analysis Model with Attribute Groups: Focused on Music Streaming Services

Sangwon Eum, Korea University Business School, Seoul, Korea, Republic of, Hosun Rhim, Jeunghyun Kim, EunSu Lee

We develop a conjoint model based on attribute groups to effectively consider massive numbers of attributes. Previous research often faced challenges in measuring individual level utilities or required additional experiments to consider multiple attributes. A new method is developed and applied to music streaming services. Attributes are identified by text mining.

2 - Understanding the Concept of Proactive Quality Management and Its Profitability: The Event Study Method

Chongwoo Park, Korea University Business School, Seoul, Korea, Republic of, Hosun Rhim, Kim Yejee

Customers tend to defect when dissatisfied with the quality of products due to inconvenient experiences. To address this issue, we propose an innovative program called 'Proactive Quality Management.' We defined the concept of proactive quality management and deployed the event study method to measure its financial performance.

3 - Video Ranking Based on Reference Effect

Mengyan Zhu, Southeast University, Nanjing, China, People's Republic of, Ningyuan Chen, Qingwei Jin

By analyzing a public dataset from NetEase Cloud Music, we empirically verify that within short-form video platforms, where viewers can swipe down to see the next video, there exists an unconscious tendency for viewers to compare new videos with the best ones they have previously watched. Based on such empirical evidence, we study the sequential product ranking problem considering this peak reference effect. Traditional video recommendation literature typically ranks videos based on their popularity. However, our study takes into account the peak reference effect and finds that platforms should provide mediocre videos first and then gradually provide better videos to manage viewers' expectation. We consider both pre-set ranking scenario and responsive ranking scenario. Although the dynamics in the model make the problem challenging, we present structural results and design tractable algorithms with guaranteed performance for both scenarios.

4 - Enhancing AI Recommendation Services for Smarter Decisions: The Role of Standardization and Customization in Optimizing User Satisfaction

Sung Yeon Kim, Korea University, Sejong, Korea, Republic of, Jaeseok Yoon, Seonghyeon Han, Jinmin Kim

This research fundamentally rethinks how AI recommendation systems are evaluated, particularly emphasizing the strategic role of standardization. By adapting traditional metrics for information and system quality, the study evaluates how these metrics influence the performance and user reception of AI-driven recommendation tools. Central to this discourse is the influence of standardization as a critical moderator that substantially enhances recommendation quality and, consequently, user satisfaction. Through the application of structural equation modeling, this analysis highlights the significant benefits of aligning system outputs with standardized quality benchmarks, which in turn fosters enhanced user engagement. Furthermore, the research emphasizes the need for maintaining an optimal balance between standardization and customization. This balance is crucial for providing personalized recommendations that are not only impactful but also drive smarter decision-making processes. Overall, the findings promote a nuanced approach to the design and implementation of AI systems, advocating for standardized frameworks that enhance both the utility and acceptability of AI recommendations in various contexts.

TB49

Summit - 441

Emerging Topics in Platform Operations and Retail

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Dmitry Mitrofanov, Boston College, Chestnut Hill, MA, United States

1 - Wage Fairness in the Gig Economy

Jingyuan Wan, University of Michigan, Ann Arbor, MI, United States, Saif Benjaafar

We consider the impact of imposing fairness constraints on the wages that an on-demand service platform pays its workers. The platform draws on a pool of independent contractors (workers) with heterogeneous opportunity costs from one or more distinct groups. The workers decide whether or not to work for the platform and how much. The platform decides on the nominal wage it pays the workers, knowing that they are sensitive to both the nominal wage and how busy they expect to be, and the price it charges customers, knowing that they are sensitive to both the price and the delay they expect to experience. The platform can also choose to cap worker participation. We consider two notions of fairness: within-group fairness (in the form of a wage floor) and across-group fairness (in the form of a constraint on the wage difference between groups). In each case, we examine the impact of the fairness constraint on labor, consumer, and social welfare.

2 - Privacy-Preserving Data-Driven Inventory Management

Lorraine Yuan, Cornell Tech/Cornell University, New York, NY, United States, Elena Belavina, Karan Girotra

The use of customer data (demographics, past purchases, etc.) for inventory management can improve firm profits and customer service levels. Yet, large-scale use of such data increases the risks of breaching customers' privacy. This paper develops differential-privacy based privacy-preserving adaptations for three data-driven newsvendor pipelines: the usual two-step predict-then-optimize method, and two newer joint predict-and-optimize methods. We characterize the tradeoff among privacy loss, profits and consumer surplus for the privacy preserving versions of these three approaches. We show that both joint approaches outperform the two-step approach to preserve privacy: Obfuscating customer data can be done with less profit loss by considering downstream optimization problems, we can privatize outputs with less noise and more targeted noise injection.

3 - Dynamic Discrimination on Online Platforms

Peibo Zhang, Emory University , Atlanta, GA, United States, Xinyu Cao, Dennis Zhang, Ruomeng Cui

Prior studies on discrimination have revealed widespread racial discrimination on online platforms. In this paper, we examine how racial discrimination evolves over time. Using observational data from a large online education platform with more than 100,000 teachers in the United States, we tracked the daily openings and bookings of teachers during their first 12 weeks after joining the platform. We observed that African American teachers initially had more daily openings than White teachers but continued to receive fewer daily bookings. Utilizing a fixed-effect model, we show that the gap in daily bookings between White teachers and African American teachers increases by 6.38 times over the 12 weeks, equating to a 22% weekly increase. We further identified that the early accumulation of a repeated customer base and the reputation effect are the two key mechanisms driving this increase in discrimination. Additionally, we found no evidence that White teachers have a faster improvement in class quality compared to African American teachers. Our results underscore that past studies focusing solely on static discrimination may underestimate the true extent of discrimination over time.

4 - Gender Biases in Customer Satisfaction Surveys: Evidence from the Customer Service Center of An Online Retail Platform

Xiao Ni, Fudan Univerisity, Shanghai, China, People's Republic of, Qiuping Yu, Tianjun Feng, Yitong Wang

Using a large-scale data set of over 10 million interactions and about 1.5 million customer satisfaction surveys from an online retail platform in China, we study customers’ gender biases against workers and the mechanisms that drive such biases. We use the routing policy of the contact center as an exogenous shock, which allows us not only to causally identify customers’ gender bias against workers but also the mechanisms. In contrast to the literature which generally shows customers’ biases against male workers in female-dominated industries, our results show that customers are biased against female workers despite them being the majority among contact center workers. Specifically, such biases leads to 1.3% lower in customer ratings for female workers compared to male workers. The biases are more severe among customers from rural areas compared to those from big cities, which is consistent with status-based discrimination.

TB50

Summit - 442

Organ Transplantation 2

Invited Session

MSOM: Healthcare

Chair: Diwakar Gupta, University of Texas, Austin, TX, United States

1 - Are Faster Out-of-Sequence Utilization Decisions Beneficial?

Jingyao Huang, University of Missouri-Kansas City, Kansas City, MO, United States, Paola Martin, Diwakar Gupta

In the US, Organ Procurement Organizations (OPOs) procure deceased-donor kidneys, whereas Transplant Programs (TxPs) make utilization decisions for candidates listed at their centers. OPOs typically make offers for several candidates at the same time, which could be listed at multiple centers, in the strict priority sequence determined by national allocation rules. Higher-ranked candidates have the right of first refusal over lower-ranked candidates. TxPs are expected to utilize kidneys based on the merits of performing a transplant for each candidate independently of other candidates. However, they frequently utilize offered kidneys for lower-ranked candidates. This phenomenon is called list diving. The prevalence and impact of list diving on recipients' survival has been studied in the literature. However, in these papers, the speed with which a TxP made its utilization decision is not considered a factor. Based on evidence from the data, we first propose a time-based rule to identify a subset of utilization decisions that were out of sequence, and then evaluate their impact on the recipient and graft survival. The out-of-sequence transplants identified via the speed-based criterion are referred to as targeted placements (TPs). Thus, this study serves to show that quick out-of-sequence utilization decisions are neither better nor worse than all other utilization decisions. Similar to list diving, TPs shorten the time on dialysis for TP recipients while lengthening it for skipped candidates. However, contrary to our prior, they have no significant effect on the 1-year post-transplant patient and graft survival among either group of candidates.

2 - Pay-for-Performance Incentives in Organ Transplantation

Saumya Sinha, University of Minnesota, Minneapolis, MN, United States, Taewoo Lee, David Mildebrath, Andrew Schaefer, A Osama Gaber, Howard J Huang

Transplant programs in the US are subject to outcome-based regulations, whereby federal agencies evaluate them based on their patients’ post-transplant survival. Clinical evidence indicates that such regulations induce programs to reject high-risk transplant candidates to avoid penalization. We present a principal-agent model to analyze the incentives created by such regulations. We show that past regulations might have induced risk-averse patient selection. As an alternative, we propose a pay-for-performance reimbursement scheme that not only penalizes programs with below-average outcomes, but also pays a bonus to those with above-average outcomes. We demonstrate that the proposed scheme incentivizes programs to improve post-transplant outcomes, without inducing adverse patient selection.

3 - Optimization Models for Kidney Paired Donation (Kpd) with Desensitization Option

Michael Fu, University of Maryland-College Park, College Park, MD, United States, Xingyu Ren, Steven Marcus

Kidney transplantation is the definitive gold standard treatment for providing the best quality of life for end-stage renal disease patients, but there is a drastic shortage of kidney donors, both deceased and living. Furthermore, even for those patients with willing living donors, certain incompatibilities (e.g., blood type) may preclude transplantation. Two approaches for expanding access to transplantation are: (i) kidney paired donation (KPD), where transplant candidates swap their incompatible donor for a compatible donor; and (ii) desensitization therapy, where transplant recipients have their antibodies treated/removed. Currently, these schemes are predominantly administered in living-donor transplants and independently of each other, and the combinatorially hard KPD optimization is addressed using static deterministic models. A multi-university collaborative team of researchers from George Mason, Maryland, and Louisville are working on an integrated approach incorporating various optimization models. This talk will overview the entire project and then focus on one part: the individual organ acceptance decision.

4 - Optimal Desensitization Therapies for Kidney-Pairing Donations

Marwan Shams Eddin, George Mason University, Euless, TX, United States, Hadi El-Amine, Naoru Koizumi, Meng-Hao Li

Kidney transplantation stands as the quintessential treatment for end-stage renal disease, providing a superior quality of life for patients. However, accessibility remains a challenge due to the shortage of donors and issues of compatibility, particularly with ABO and HLA types. Current strategies such as kidney paired donation (KPD) and desensitization therapy address these challenges independently but face limitations. The decisions of pairing donors with recipients are fundamentally driven by the probability of recipients developing future diseases, (e.g., graft failures, cancer, among others) under specific treatments. In this work, we build a customized KNN machine learning model coupled with an optimization-based framework to determine desensitization treatments that reduce risks of undesired outcomes. Our results indicate that the customized treatments can reduce the risk of such outcomes.

TB51

Summit - 443

Technology and Operations

Invited Session

MSOM: Service Operations

Chair: Dennis Zhang, Washington University in St Louis, ST LOUIS, MO, United States

Co-Chair: Xiaoyang Long, University of Wisconsin-Madison, Madison, WI, United States

1 - Operational Strategies for Customer Service: A Gatekeeper Framework

Maqbool Dada, Johns Hopkins University, Baltimore, MD, United States, Brett Hathaway, Evgeny Kagan

Customer service has evolved beyond in-person visits and phone calls to include live chat, AI chatbots and social media, among other contact options. Service providers typically refer to these contact modalities as “channels”. Within each channel, customer service agents are tasked with managing and resolving a stream of inbound service requests. Each request involves milestones where the agent must decide whether to keep assisting the customer or to transfer them to a more skilled – and often costlier – provider. To understand how this request resolution process should be managed, we develop a model in which each channel is represented as a gatekeeper system. We characterize the structure of the optimal gatekeeping policy and identify conditions under which threshold policies yield the optimal solution. We then turn to the broader question of strategic selection of an appropriate mix of service channels. We solve the channel mix problem for the special case where the service provider chooses among three channel architectures: live-agent-only, chatbot-only or both, and show that, in equilibrium, all three architectures may emerge as optimal.

2 - The Impacts of Recommendations on Consumption and Creation on Online Content-Sharing Platforms

Zhiyu Zeng, Washington University in St. Louis, St. Louis, MO, United States, Zhiqi Zhang, Dennis Zhang, Tat Chan

Online content-sharing platforms such as TikTok and Facebook, heavily rely on advanced algorithms to recommend user-generated content (UGC) to their users. These users can simultaneously consume and create UGC. Previous research has shown the content recommended by algorithms significantly impacts users' consumption habits; however, these recommendations can also affect users' motivation to create content, since they provide insights into the works of other content creators and generate peer effects. To study the impacts of recommendations on both consumption and production, we conducted a randomized field experiment on one of the world's largest online video-sharing platforms. In this experiment, we manipulated the popularity levels of the videos recommended by the algorithm. Our findings demonstrate that when users encounter recommended videos with lower popularity, their video-viewing time decreased by 1.38%. However, we also observed a substantial 2.39% increase in the number of videos uploaded to the platform. To understand and explain these results, we utilized a structural model that allows a user's incentive to create videos to be influenced by the "peer effect" captured by the relative popularity of recommended video creators and the user. We then used the estimation results of the structural model to conduct counterfactuals. Results reveal that recommending videos with high popularity is not always optimal when both content consumption and production are crucial to the platform. To maximize overall value, the algorithm should be tailored to recommend videos of varying popularity levels, better aligning with individual users' preferences for consumption and production.

3 - The Impact of Foster Care Payments on Child Placement Stability

Wani Zhang, London Business School, London, United Kingdom, Kamalini Ramdas

We examine the effect of monthly foster care payments on the placement stability of foster care children using data from the Adoption and Foster Care Analysis and Reporting System (AFCARS) during 2008-2021. We leverage the longitudinal nature of the dataset to assess how variations in payment levels affect placement stability, controlling for child-specific factors and employing an instrumental variable strategy to address potential endogeneity issues. We will present the main effects of the treatment as well as differential treatment effects, and we will discuss policy implications.

4 - The Granularity of Wait TIME Information: A Large-Scale Field Experiment on a Major Ride-Sharing Platform

Yiming Zhang, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Qiuping Yu, Yong-Pin Zhou

Collaborating with a ride-sharing platform, we study whether and how the granularity of wait time information (WTI) impacts customers abandonment behavior through a randomized field experiment on our partner platform. In the experiment, we consider a point estimate, a narrow interval, and a wide interval as WTI. To uncover the fundamental mechanism, we propose a structural model to explore the impacts of WTI granularity on customers' prior beliefs and waiting cost-reward ratios. Through counterfactual studies, we explore how the design of WTI granularity impacts the performance of virtual queues. Additionally, we discuss the managerial implications of our findings.

5 - On state- and time-dependent service processes in healthcare

Hao Ding, Auburn University, Auburn, AL, United States, Sokol Tushe, Ruomeng Cui, Donald Lee

The empirical operations literature has shown that the service process of customers is state-dependent, with a primary focus on the relationship between patient length-of-stay (LOS) and workload in healthcare settings. However, the service process is also time-dependent because physician workload is not a static state, but varies continuously over the course of a patient's stay. By recognizing that the service process is both state- and time-dependent, we argue that the instantaneous service rate is the more natural quantity of interest in place of LOS. Conducting analyses based on the instantaneous service rate rather than LOS allows us to potentially resolve a longstanding question regarding the relationship between service speed and workload. Specifically, the empirical literature has identified contradictory findings among analyses based on LOS. We show, through the setting of an academic emergency department, that a consistent and intuitive relationship emerges when the analysis is based on the instantaneous service rate. We also use a naturalistic simulation to demonstrate meaningful performance gains from employing a service rate model to set staffing, relative to staffing according to results from analyses based on LOS.

TB52

Summit - 444

Research on Innovative Service Operations

Invited Session

MSOM: Service Operations

Chair: Ming Hu, Rotman School of Management, University of Toronto, Toronto, ON, M5S 3E6, Canada

Co-Chair: Zhoupeng (Jack) Zhang, Rotman School of Management, University of Toronto, Toronto, Canada

1 - Phantom Capacity: Informational Cost of Multihoming on Ride-hailing Platforms

Kaitlin Daniels, Washington University in St. Louis, Olin Business School, Saint Louis, MO, United States, Jaelynn Oh

We study a ride-hailing market in which two platforms compete and drivers multihome. This environment produces information asymmetry that drivers can strategically exploit. When a driver is busy completing a ride for platform A, platform B does not observe that the driver is busy. The driver may strategically appear available to platform B (e.g. by logging into platform B's app) to increase the driver's utilization. We refer to drivers deploying this tactic as "phantom drivers." We derive phantom drivers' optimal strategy and quantify the inefficiency imposed on matching. We then compare to the alternative where platforms enforce single homing and describe the conditions under which platforms are willing to sacrifice the pooling benefits of multihoming to avoid the informational cost of phantom drivers.

2 - Solving Wild Goose Chase: Spatial, Temporal Pooling, and Service Region Reduction

Mingliu Chen, The University of Texas at Dallas, Richardson, TX, United States, Ming Hu

Wild Goose Chase is a phenomenon that may greatly hinder efficiencies in the ride-hailing context. A driver may drive a significant time for a pick-up but end up with a short ride. We propose implementing spatial, temporal pooling, and service region reduction to improve operational efficiencies in such systems. In spatial pooling, a driver only picks up riders at certain locations instead of chasing them around the city. However, riders must first walk to the pick-up location before getting served. We show that the pooling mechanisms may greatly improve operational efficiencies. The benefits critically depends on the rider's disutility of traveling on foot, the service radius, and whether the platform can endogenize the demand. With temporal pooling, a driver may not serve riders with the first-come-first-serve discipline. Instead, a driver can wait until certain riders are accumulated before serving them. By doing so, the expected pick-up distance for the closest rider decreases when the number of riders increases. Finally, the service region can be reduced to prevent long pick-up distances. However, we show that there exists a threshold and only above which, service region reduction is beneficial.

3 - No Call, No Show: Impact of No Show and Its Remedies on Service Platforms

Nil Karacaoglu, The Ohio State University, Columbus, OH, United States, Simin Li

With the surge in online-to-offline services, platforms face growing concerns about service provider no-shows, such as a handyman hired on Handy not showing up or a delivery person missing a mutually agreed time window. Customers may return to the platform after their own no-shows, but platforms risk losing those who experience service provider no-shows. Addressing no-show incidents is paramount for these platforms. Analyzing data from a European cleaning platform, we empirically assess how service provider no-shows impact customer retention. Our study shows that service provider no-shows increase the probability of customer attrition by 33.6%, with the negative impacts being more pronounced for more loyal customers. We further utilize a customer lifetime value model and show that the platform experiences a 4.2% loss in customer lifetime value due to no-shows. Subsequently, we study the effectiveness of two remedies for no-shows. We find no association between same-day rescheduling and no-show recovery, while offering a discount is positively linked to alleviating customer

attrition following no-shows. Lastly, we develop a model-based framework to suggest optimal remedy strategies, including offering discounts and maintaining a list of flexible standby cleaners to mitigate the negative impact of no-shows.

4 - Riding Through Rallies: Will You Tip More?

Wanjiang Deng, School of Management, Zhejiang University, Hangzhou, China, People's Republic of, Zhoupeng (Jack) Zhang, Ming Hu

Other than gig platforms' interventions, what "organic" factors can make customers treat gig workers (e.g., Uber drivers) in a more caring manner? In this paper, we approach this question by examining the impacts of the 2020 Chicago George Floyd protests on passengers' tipping behavior in the ride-hailing marketplace. We collect the ride-hailing data and deploy a year-to-year difference-in-differences analysis. We find that the protests not only increase the probability of passengers leaving a tip but also elevate the tip ratio by 2.98 percentage points. Nevertheless, this positive impact is attenuated on trips with less exposure to protest events and ebbs away two to three days after the protests concluded; it increases gig workers' incomes by a yet modest amount. By comparing with the taxi market, we identify a relatively weak social norm of tipping in the ride-hailing marketplace as one critical driving force: on the one hand, the weak norm makes passengers tip less generously and much less often in a ride-hailing market than they would in a taxi market; on the other hand, though, it also leaves space for event-specific and market-based factors to weigh in and (at least temporarily) affect passengers' behaviors. We propose two such factors that are potentially at play during the protests: passengers' empathy and platforms' disruption-aware services. Our work reveals how social movements may reshape customers' behaviors and, furthermore, shed light on the broader impacts of tipping as well as the platforms' role in fostering a tipping norm in the on-demand economy.

TB53

Summit - 445

Topics in Sustainability and Food Supply Chain

Invited Session

MSOM: Sustainable Operations

Chair: Tharanga Rajapakshe, University of Florida, Gainesville, FL, 32608, United States

Co-Chair: Na Zhang, Warrington College of Business, University of Florida, Gainesville, FL, United States

1 - Production Rationing Across Different Regulation Clusters: Carbon Emissions Perspective

Avinash Geda, University of North Carolina Wilmington, Wilmington, NC, United States, Nazli Turken

Supply chains still remain largely global, even after the vulnerabilities exposed after the recent disruptions. Lower costs of production and conducive regulation environments are the major contributing drivers. In the context of global supply chains, production could also get rationed across different geographical regions. From a carbon emissions perspective, different clusters could be subjected to different regulations, which directly plays into the production rationing decision. In this paper, we look at a firm's production rationing decisions across different geographical regulation clusters. When analyzing a firm's production rationing decisions within different regulation clusters, our focus is on understanding the influence of symmetric and asymmetric regulations on these decisions across two clusters. Symmetric regulations entail the same carbon emission regulation applied across clusters, while asymmetric regulations involve differing regulations in each cluster. By examining how firms allocate production resources in response to varying regulatory environments, we derive insights into the equilibrium production decisions. This research aims to uncover the underworking of production rationing strategies under different regulatory scenarios, providing insights for both policymakers and businesses.

2 - Non-Profit Support in Education: Resource Allocation and Students&Rsquo; Lifetime Outcomes

Harish Guda, Arizona State University, Tempe, AZ, United States, Goutham Takasi, Milind Dawande, Ganesh Janakiraman

An important determinant of the effectiveness of a non-profit organization is its resource-allocation strategy, i.e., how it assigns resources to its beneficiaries. Our work is motivated by resource allocation among nonprofit initiatives in which the outcome to a beneficiary depends on their effort, such as in education. In particular, we focus on nonprofit initiatives -- supported by an NPO -- that adopt a two-stage structure in allocating resources to its beneficiaries. The first-stage support is standardized and is provided to all beneficiaries, while in the second stage, the quality of support provided to the beneficiaries is contingent on their first-stage performance. Specifically, we assume that the NPO can provide a beneficiary with either a base-quality resource or a superior resource in the second-stage. The lifetime outcome of a beneficiary, a key component of the NPO's objective, is increasing in the beneficiary's effort in both the stages. Further, providing the beneficiary with the superior resource in the second-stage complements their effort and results in a better lifetime outcome as compared to providing the base-quality resource. We analyze the strategic role of an NPO's resource allocation strategy on the agent's effort and their lifetime outcome. We show why an NPO benefits from deliberately restricting access to resources, even in the absence of resource scarcity. Finally, we analyze the effect of resource scarcity and competition among the beneficiaries on their effort and lifetime outcomes.

3 - Strategies for Fair Food Distribution: Overcoming Uncertainty in Donations and Food Waste

Xianjia Gao, University of Illinois Urbana-Champaign Champaign, Champaign, IL, United States

Food bank, a non-profit humanitarian organization, secures charitable food donations through collaborative partnerships. Food banks are pivotal in mitigating food insecurity by distributing donations to those in need. However, they face several significant challenges. First, food banks depend on uncertain donations from diverse sources such as supermarkets, farms, food manufacturers, and individuals, leading to a highly variable supply. Second, the demand from partner agencies varies, complicating the equitable distribution of food. Moreover, approximately 35% of the distributed food (equating to about 29 million pounds annually) is perishable, making the management of food storage and distribution more complex. To address these challenges, we develop a multi-period dynamic model to help food banks allocate perishable food to agencies, aiming to maximize the value of the food, achieve equitable distribution, and minimize waste. We propose a heuristic algorithm based on fluid approximation with a provable regret upper bound. Furthermore, to tackle the dynamic environment, we further refine the heuristic through reoptimization and prove that the same regret bound can be achieved. Our results demonstrate that the proposed heuristics achieve asymptotic optimality with respect to the size of the food distribution network.

4 - Retailing Strategies of Imperfect Produce and the Battle Against Food Waste

Haoran Yu, Syracuse University, Syracuse, NY, United States, Burak Kazaz, Fasheng Xu

Edible but not-quite-perfect produce is often not sold in stores, leading to substantial food waste. The commercialization of imperfect produce constitutes a promising lever to tackle food waste. This study investigates how a grocery retailer should choose from three popular retailing strategies: discarding imperfect produce, bunching with cosmetically-perfect produce, and differentiating by selling perfect and imperfect produce separately at different prices. We adopt the multinomial logit choice model to derive consumers' demand and determine optimal decisions under each strategy. We compare three strategies and provide managerial insights to retailers about when each strategy should be used. We further examine two common policy interventions for reducing imperfect produce waste: relaxing grading standards to allow more imperfect produce to get into stores; and educating consumers on the value of imperfect produce. Our results suggest that policymakers should exercise caution in implementing these interventions because relaxing grading standards may shift food waste from farms to retail stores when the discarding strategy is optimal; and educating consumers about the value of imperfect produce can boost its consumption but may also lead to reduced sales of perfect produce due to cannibalization, potentially causing the retailer's shift to the discarding strategy and contributing to increased food waste. Our study also presents four extensions: upcycling the imperfect produce into by-products, endogenous pricing of imperfect produce, the mixed strategy that combines multiple popular retailing strategies, and the full-shelf ordering policies. The analyses of these extensions confirm the robustness of our main findings.

5 - Production Planning for Multiple Products Incorporating Emissions Trading

Ziyan Zhou, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Tong Wang, Jingwei Zhang

We consider a finite-horizon dynamic control problem where the manufacturer produces N products and satisfies random demands under a given total emission allowance level over the planning horizon. Meanwhile, the firm engages in managing the emission allowance by participating in the carbon trading market through transactions. We developed a heuristic policy that integrates emissions trading and production to minimize total cost, ensuring that its performance is within O(N^(1/2)) of the optimal solution.

TB54

Summit - 446

Diverse Topics on Responsible Operations Management

Invited Session

MSOM: Sustainable Operations

Chair: Mahyar Eftekhar, Arizona State University, Tempe, AZ, United States

Co-Chair: Vanitha Virudachalam, Gies College of Business, UIUC, Champaign, IL, United States

1 - Behavioral Analysis of Needs Reporting and Fund Allocation in Humanitarian Response

Bengisu Urlu, INSEAD, Fontainebleau, France, Karen Donohue

When a disaster strikes, an accurate assessment of required funds for the response is crucial to utilize scarce resources for the greatest good. Humanitarian organizations (HOs) evaluate the needs and report the required amount of funds for their response operations. Donors then allocate their budget informed by these requirements. We represent this setting through a game theory model and explore how the type of information that the donor receives (e.g., targets for the number of people being served versus the cost of service per person) influences their allocations across HOs through experimental studies. We also examine how HOs form their reports under different information types, and the extent of their trustworthy behavior. Finally, we analyze the impact of these dynamics on humanitarian outcomes.

2 - The Production Benefits of Personalized Agricultural Advice: a Case Study of Mkrishi in Buland Shahar

Sriram Venkataraman, University of South Carolina, Columbia, SC, United States, Campbell Clarkson, Necati Tereyagoglu

We investigate the production effects of introducing mobile phone-based agricultural advice to farmers, using the launch of the mKrishi platform in Buland Shahar district in India. We find that, after a pilot of the mKrishi platform was launched in 2008, potato production in Buland Shahar increased by 3.93% relative to a synthetic control district. We then explore if such an effect could be explained by the changes in farmers' actions in three key agricultural inputs: cultivated area, chemical fertilizer usage, and irrigated area. We also show aggregate-level evidence of an increase in mobile phone usage in Buland Shahar after the launch of mKrishi, relative to the synthetic control, thereby suggesting the use of mKrishi by the farmers.

3 - Integrating Life Cycle Assessment into Supply Chain Optimization

Selin Hülagü, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, Wout Dullaert, Ayse Sena Eruguz, Reinout Heijungs, Dirk Inghels

Companies face increasing pressure from both government regulations and consumer demands to enhance the environmental sustainability of their supply chains. Life Cycle Assessment (LCA) appears to be one of the most promising approaches to systematically assess the environmental impacts of supply chains. LCA is typically performed for products for which decisions are already taken. Decisions on the network structure to bring products and/or services to customers are the typical focal point of supply chain network optimization (SCO). Existing literature handles LCA and SCO as separate sequential steps, leading to inconsistencies in scope and challenges in data transfer and rescaling. This research introduces a novel Supply Chain Life Cycle Optimization (SCLCO) model that integrates LCA and SCO. Our SCLCO model is based on LCA data structure and encompasses multi-time closed-loop SCO decisions (e.g. reverse chain management, inventory control, network design) while considering the three pillars of sustainability; environmental, economic, and social. It includes the harmonization of principles, terminology, and notation thereby bridging the gap between the LCA and SCO communities through a generalized formulation. Computational experiments on a selected SCO model from Operations Research literature validate the SCLCO, demonstrating its effectiveness in providing valuable insights to both LCA and SCO practitioners and researchers. Results emphasize that the simultaneous execution of LCA and SCO in SCLCO minimizes the risk of overlooking decision impacts and facilitates data transfer from existing LCA databases.

4 - Too Good to Go: Combating Food Waste with Surprise Clearance

Luyi Yang, University of California, Berkeley, Berkeley, CA, United States, Man Yu

We study a novel clearance mechanism of perishable goods to reduce food waste and increase store profit.

TB55

Summit - 447

Analytics for the Detection and Disruption of Illicit Supply Chains

Invited Session

Public Sector OR

Chair: Margret Bjarnadottir, University of Maryland, College Park, MD, United States

1 - Improved Lisa Analysis for Zero-Heavy Crack Cocaine Seizure Data

Eunseong Jang, University of Maryland, College Park, MD, United States, Margrét Bjarnadóttir, Marcus Boyd, S. Raghavan

Local Indicators of Spatial Association (LISA) analysis is a useful tool for analyzing and extracting meaningful insights from geographic data. It provides informative statistical analysis that highlights areas of high and low activity. However, applying LISA analysis methods to zero-heavy data may not be appropriate, as without the correct mathematical context the meaning of the patterns identified by the analysis may be distorted. We highlight these issues through statistical analysis and provide the appropriate context for interpreting LISA results for zero-heavy data. We then propose an improved LISA analysis method for spatial data with majority zero values. Applying our proposed methodology to crack cocaine seizure data in the U.S., we demonstrate its benefits in zero-heavy contexts as well as finer categorization of observed data. As LISA analysis is a popular approach that supports policy analysis and design, and as zero-heavy data is common in these scenarios, our work provides a framework that is better tailored for result interpretation and an improved approach for zero-heavy geographic data, ultimately leading to better decisions.

2 - Large Language Models for Scalable Text Data Extraction and Analysis in the Context of Drug-Related Violence in Colombia

Shriyan Reyya, Montgomery Blair High School, Silver Spring, MD, United States, Margret Bjarnadottir, David Anderson, Galia Benitez

Extracting actionable information from large text corpora at scale has immense potential, ranging from improving business predictions and decisions to revolutionizing social science research. Historically, however, the cost of manual data collection and extraction has hindered the use of such databases. Large language models hold great promise in automating and accelerating this process. We demonstrate this potential, utilizing GPT-3.5 to extract information from a public dataset of 280,000 Colombian news articles focused on violent events occurring over four decades. After deduplication and verification, our dataset is more comprehensive than any other available data source, opening up a range of research possibilities. For example, we have studied the impact of forced coca crop eradication on violence, finding no significant impact, which conflicts with literature that has based its analysis on smaller data. Moreover, we have utilized the data for violence predictions and the study of criminal groups and their territories. The impact of our research is two-fold: firstly, we demonstrate the capability of large language models in scalable text data extraction; secondly, in the context of Colombia, our data and results challenge current literature and stand to assist in informing public policy.

3 - Analyzing Illicit Drug Supply Chains: A Cocaine Case Study

S. Raghavan, University of Maryland, College Park, MD, United States, Galia Benitez, Margret Bjarnadottir, Siddharth Chandra, Greg Midgette

We report on an interdisciplinary effort of analyzing illicit drug supply chains. Our work has focused on illuminating the structure of illicit drug supply chain networks, quantifying the impact of disruptions on the drug supply chain and building models to support disruption strategies of anti-narcotics and other agencies. In this talk we will summarize some of our key findings and approaches.

4 - Analytical Methods to Identify Broad Patterns in the Commercial Sex Ad Ecosystem

roya shomali, The University of Alabama, Tuscaloosa, AL, United States, Nickolas Freeman

Sex traffickers use the internet to exploit individuals through online commercial sex advertisements (sex ads). Analyzing sex ad data to derive actionable insights poses significant challenges due to the high volumes of ads, the prevalence of scam ads, and the presence of generic or ambiguous data points that introduce noise. Consequently, there is a limited understanding of the individuals represented in these ads and the hidden patterns they follow. This research develops and analyzes a comprehensive set of features to uncover broad categories of individuals engaged in the online commercial sex ecosystem. Our methodology integrates techniques from natural language processing and network science to analyze individual behaviors in sex ad data, focusing on movement patterns, linguistic choices, and posting behaviors. We employ a binary classification framework to estimate the similarity between individuals, thereby quantifying their degree of similarities. We then use graph-based community detection methods to identify distinct clusters of individuals within the online ecosystem with high similarity among individuals in the same cluster. Our goal is to uncover insights regarding the complex patterns obscured within the data that can support counter-trafficking agencies by revealing how different individuals or groups, potentially across various regions, operate within this illicit network. This research advances academic discourse in trafficking analysis and provides scalable, practical tools that meet the diverse needs of real-world applications. These tools offer crucial support to law enforcement and policymakers, enabling more informed strategies to combat sex trafficking.

TB56

Summit - 448

Optimization for Real People

Invited Session

Minority Issues Forum

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, 80401, United States

1 - Comparative Performance of Battery- and Diesel-Powered Haulage Fleets in Underground Mines

Aaron Swift, Colorado School of Mines, Golden, CO, United States

Battery-electric vehicles have the potential to improve ambient conditions in underground mines by reducing the amount of heat and diesel exhaust emitted. Existing literature considers the differences in cost, heat load, and productivity between individual diesel and battery vehicles; this study uses simulation techniques to determine the effect of electrifying an entire haulage fleet on the overall heat load, emissions, and energy usage of an underground mine as activities take place deeper in the earth. This information is important for mine operators to assess vehicle electrification as a means to achieve greenhouse gas emissions reduction targets.

2 - Optimizing Vehicle Fleet Composition in Underground Mines

John Ayaburi, Colorado School of Mines, Golden, CO, United States, Aaron Swift, Jason Porter , Andrea Brickey , Alexandra Newman

Underground mine planning uses production schedules to determine a (near-)optimal sequence of activity execution to maximize net present value while considering resource limitations and spatial precedence. At the time of this writing, the mining industry relies heavily on the use of diesel-powered equipment, which accounts for heat accumulation and exhaust emissions that can create unsafe conditions in the work environment. Sustainable mining practices call for the transition from diesel- to battery-powered equipment. We present a large-scale production scheduling model that (i) prescribes activity start times in a medium-term schedule at daily fidelity, taking into account ventilation and refrigeration; and, (ii) determines a fleet composition, relative to a diesel-only fleet, that improves productivity. We implement an enumeration technique, embedded in our optimization model, whose special structure we exploit, to solve otherwise-intractable medium-term schedules to within 10% of optimality with solution times ranging from 145 seconds - 173,000 seconds. We find that the need for refrigeration is delayed and exhaust emission is reduced as more battery-powered equipment is introduced, showcasing the utility of battery vehicles in maintaining productivity and improving the safety of underground work environments. Our results provide schedules and fleet composition(s) that enhance the integration of battery vehicles into an operation. Our model provides a tool with which mine planners can make environmentally friendlier strategic decisions.

3 - Addressing School Accessibility Disparities in New York City: An Application of Vulnerability-Based Perspectives Through Computational Urbanism

Jose Ramirez-Marquez, Stevens Institute of Technology, Hoboken, NJ, United States, Rosario Contreras

This study investigates communities' vulnerability regarding school accessibility in New York City. Based on methods from systems reliability and computational urbanism, we analyze the spatial distribution of schools and their accessibility across different communities in the city. Our research aims to uncover significant disparities in school access, shedding light on potential inequalities in educational opportunities among diverse neighborhoods. By examining factors such as proximity, transportation infrastructure, and socioeconomic indicators, we provide insights into the spatial dynamics of educational accessibility. Our findings underscore the importance of addressing reliability, vulnerability, and resilience issues within urban contexts, particularly in critical areas such as education. Furthermore, we advocate for the wider adoption and integration of computational urbanism approaches within the research community to comprehensively understand and mitigate socio-spatial disparities, thus promoting the development of more equitable and resilient urban environments.

4 - Assessing Overall Performance of spORts Clubs and Decomposing into on-Field and off-Field Efficiency: An Application to Australian Rules Football

Don Galagedera, Monash University, MELBOURNE, Australia

Generally playing group management performance and financial management performance of sports clubs are assessed separately. We adopt a non-parametric methodology to assess overall performance first conceptualising overall management as a production process comprising of two serially-linked sub-processes; playing group management and financial management. Thereafter, we decompose overall performance to obtain estimates of performance at the sub-process level. Through this procedure it is possible to determine whether it is a sports club’s on-field performance or off-field performance or both may contribute towards its inefficiency, if any, in overall management. Further, a model is developed to determine targets for inefficient clubs to become overall efficient. The method is applied to 18 clubs in Australian rules football league. In 2021 season, the results reveal that on-field performance, on average, is better than off-field performance and variability in off-field performance is higher than that of on-field performance. Observed overall management inefficiency is mainly due to inefficiency in financial management. Results are robust to weighting scheme adopted in overall efficiency decomposition.

5 - Pilatus Aircraft Optimizes Inventory Management

Alexandra Newman, Colorado School of Mines, Golden, CO, United States, Garrett Miyaoka, Tulay Flamand

Pilatus Aircraft Ltd markets, sells, and maintains all Pilatus aircraft in North and South America, and seeks to improve customer service levels of all its repair parts, with an emphasis on "critical parts" orders associated with an aircraft-on-ground status. Using a hybrid methodology that combines a (Q,R) inventory stocking model with an optimization model that recommends antiquated parts to liquidate from its inventory, our tool improves the mix of parts held at the Broomfield warehouse to more quickly service customer orders. We describe actions that the company has taken to improve its service level factor, reduce its demand variability, and stabilize lead times across its supply chain for the sake of customer service.

TB57

Summit - Terrace Suite 1

Optimization and AI for Precision Health of Dynamic Diseases

Invited Session

Health Applications Society

Chair: Anil Aswani, UC Berkeley, Berkeley, CA, United States

Co-Chair: Joo Seung Lee, University of California, Berkeley, Emeryville, 94720, United States

1 - Optimizing Scheduling in Multi-Service Systems with Policy-Design Constraints: Application to University Mental Health Services

Hrayer Aprahamian, Texas A&M University, College Station, TX, United States

We present an optimization-based methodology for identifying time-varying scheduling policies, considering multiple customer classes with non-stationary arrival processes and realistic policy-design constraints. Initially, we analyze a simplified setting involving a single service type and introduce a novel approximation scheme to estimate the transient behavior of the stochastic system. By utilizing this scheme, we formulate the problem as a challenging mixed-integer nonlinear program. To address this problem, we employ a decomposition scheme to divide it into two more manageable subproblems, each of which presents a mixed-integer nonlinear program. By establishing key structural properties of the subproblems, we develop polynomial-time schemes that solve each of them to global optimality by casting them as more tractable shortest path problems. We then extend the solution scheme to accommodate multiple service types using a branch-and-bound-based method. To illustrate the effectiveness of our methodology, we perform a case study on college counseling centers. Our approach identifies scheduling policies that enhance students' access to vital mental health services by up to 52% over current practices, while utilizing the level of resources.

2 - Generating Data-Driven Clinical Role Models for Personalized Treatment Target Selection

Che-Yi Liao, Georgia Institute of Technology, Atlanta, GA, United States, Esmaeil Keyvanshokooh, Gian Garcia

Chronic diseases have a significant impact on global mortality rates and healthcare costs. Accordingly, clinicians may aim to determine treatment targets for individuals with chronic diseases to minimize the risk of adverse health events, using newly developed health tools, e.g., machine learning-based risk scoring tools. However, it is challenging to identify personalized treatment targets that are realistic, actionable, and robust to changes in the risk scoring tools by these tools alone. Additionally, persuading patients to take on the designed treatment plan can as well be difficult. To this end, we propose an approach named Data-driven Robust Clinical Role Model Selection (DR-CRS), incorporating two seemingly unrelated data sources maintained in the hospital. First, DR-CRS leverages the EHR database to find patients with similar health conditions and treatment plan. Second, it incorporates risk scoring tool records to mitigate the risk of failing to achieve treatment objectives. We develop an Active Learning Algorithm for Robust Role Model Selection (ALARRMS) and establish its exponential convergence rate. Utilizing two large and granular datasets, we illustrate the superior performance of DR-CRS compared to other benchmarks. Furthermore, we introduce two feature importance metrics to help clinicians understand how deviations from treatment targets can affect a patient's risk of adverse outcomes, thereby facilitating the design of appropriate treatment plans. Our findings highlight the effectiveness and practical relevance of DR-CRS in personalized treatment and clinical decision-making.

3 - Learning Personalized Treatment Strategies with Predictive and Prognostic Covariates

Andres Alban, Frankfurt School of Finance & Management gGmbH, Frankfurt, Germany, Stephen Chick, Spyros Zoumpoulis

We consider the problem of sequentially allocating sample observations to learn personalized treatment strategies, motivated by the design of adaptive clinical trials that aim to learn the best treatment as a function of patient covariates. In such settings, there may be clinical knowledge of which covariates are predictive (they may interact with the treatment choice) and which are prognostic (they may influence the outcome independent of treatment choice). We extend the expected value of information (EVI)/knowledge gradient (KG) framework to develop useful heuristics for a context with predictive and prognostic covariates. We show that several of our proposed allocation policies are asymptotically optimal in learning treatment strategies. When clinical knowledge is not available, we propose a Lasso method to adaptively learn which covariates are predictive and/or prognostic. Simulation experiments, motivated by applications for clinical trial design to assess potential personalized treatments of sepsis, show that our proposed policies can improve the efficiency of clinical trials to learn a treatment strategy.

4 - Using State Space Models for Personalized Lapse Prediction

Eric Pulick, University of Wisconsin - Madison, Madison, WI, United States, John Curtin, Yonatan Mintz

For individuals recovering from alcohol use disorders, lapse is an important precursor to relapse. Reliable lapse prediction is thus a prerequisite for healthcare providers seeking to effectively allocate scarce treatment resources. We consider a dataset of recovering patients who submitted once-daily ecological momentary assessments over a three-month study period. We demonstrate that state space models can be used to make effective lapse predictions compared to current population-level machine learning models.

5 - Methodology for Interpretable Reinforcement Learning for Optimizing Mechanical Ventilation

Joo Seung Lee, University of California, Berkeley, Emeryville, CA, United States, Malini Mahendra, Anil Aswani

Mechanical ventilation is a critical life-support intervention that uses a machine to deliver controlled air and oxygen to a patient's lungs, assisting or replacing spontaneous breathing. While several data-driven approaches have been proposed to optimize ventilator control strategies, they often lack interpretability and agreement with general domain knowledge. This paper proposes a methodology for interpretable reinforcement learning (RL) using decision trees for mechanical ventilation control. Using a causal, nonparametric model-based off-policy evaluation, we evaluate the policies in their ability to gain increases in SpO2 while avoiding aggressive ventilator settings which are known to cause ventilator induced lung injuries and other complications. Numerical experiments using MIMIC-III data on the stays of real patients' intensive care unit stays demonstrate that the decision tree policy outperforms the behavior cloning policy and is comparable to state-of-the-art RL policy. Future work concerns better aligning the cost function with medical objectives to generate deeper clinical insights.

TB58

Summit - Terrace Suite 2

Sequential Decision Making and Reinforcement Learning Applications in Healthcare

Invited Session

Health Applications Society

Chair: Yonatan Mintz, University of Wisconsin Madison, Madison, WI, United States

1 - Optimal Enrollment TIMES in Early-Stage Clinical Trials

Hanwen Liu, Clemson University, Clemson, SC, United States, Qi Luo, Amin Khademi

Early-stage clinical trials (CTs), including pre-Phase I and Phase I CTs, are crucial for establishing the safety profile of investigational drugs and determining dose-limiting toxicities (DLTs). This research focuses on modeling patients' decision-making processes in early-stage CTs using a Bayesian Optimal Stopping Problem framework, factoring in the efficacy of doses, the evolution of beliefs regarding dose toxicity, and the deterioration in the patient's health. Early-stage CTs adopt a dose-escalation method, where cohorts of patients are sequentially administered increasing doses of the drug until severe DLTs are observed. A control-limit policy for patient enrollment decisions was established with respect to the patient's health state and belief about the dose toxicity. Modeling patients' early-stage CT enrollment decision-making processes lays the groundwork for developing more effective and interpretable strategies for patient recruitment and retention, ultimately contributing to the acceleration of drug development processes.

2 - Data-Driven Patient Flow Optimization

Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States, Mark P. Van Oyen, Esmaeil Keyvanshokooh, Arlen Dean

Hospitals, as critical infrastructure for population health, face many difficult challenges in converting systems and patient information into evidence-based and personalized decisions. High-quality inpatient hospital care demands the effective and efficient use of hospital beds. However, capacity limitations, high variability across patients, and many human-driven processes make for a challenging environment. Advances in Industrial Engineering (IE) and Applied Operations Research promise to integrate traditional methods with advanced data analytics and decision-making methods to harness the rapidly growing Electronic Health Records (EHR) information to yield unprecedented decision support for this highly complex and essential environment. While Machine Learning and Reinforcement Learning (RL) are evolving rapidly, research from an IE perspective is essential due to the need to capture the dynamics of the system, which feature the complexities of queuing as patients move through a network of service resources, with an emphasis on resource allocation.

By focusing on the patient admission control problem, we aim to reduce adverse events (AEs) while maintaining patient safety and care quality throughout their hospital experience. To optimize a stochastic system model and minimize preventable AEs, the RL method requires a Markov Decision Process (MDP). Moreover, given serious concerns about "safe" policies that restrict online learning's potential, we should explore offline RL algorithms as a conservative approach to mitigate the potential risks associated with online learning. Therefore, to make safe and optimal sequential decisions about patient admissions, we utilize MDP models and a Safe Offline Off-Policy Reinforcement Learning algorithm.

3 - Model Based Learning and Optimization for Personalized Heparin Dosing

Qinyang He, UW-Madison, MADISON, WI, United States, Yonatan Mintz

One of the key challenges in sequential decision making is optimizing systems safely in the case of partial information especially when both states and system dynamics are unknown. For instance, the setting of computing heparin doses for patients fits this paradigm since the concentration of heparin in the patient cannot be measured directly and the rates at which patients metabolize it vary greatly between individuals. To address above challenges in heparin dosing we build a predictive model based on pharmacokinetics parameterized individually by patient. Using this personalized model, we can accurately infer patients' states and predict future therapeutic effects. With this learned information, we propose dosing algorithms that robustly keep patients safe in face of observational errors.

TB59

Summit - Ballroom 1

APS PhD Student Showcase 1

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

TB60

Summit - Ballroom 2

Supply Chain Planning and Execution

Invited Session

The Practice Section of INFORMS

Chair: Chun Ye, Amazon.com, Seattle, WA, United States

Co-Chair: Zhikun Gao, Amazon, Bellevue, WA, United States

1 - Validating Network Planning via Customer Centric Baselines

Qi Chen, Amazon, Bellevue, WA, United States

To align resourcing decisions in a large-scale supply chain network, Sales and Operations Planning generates optimal aggregate flow and capacity plans given customer demand and vendor supply projections. These plans must strike a balance between incentives of disparate business units. We evaluate common baseline-generation approaches for assessing plan and execution quality against instances of a network model. We show that our choice of methodology may lead to conflicting interpretations of plan quality.

2 - Middle Mile Capacity Planning

Ioannis Spantidakis, Amazon, Seattle, WA, United States

Amazon's middle mile network is the backbone of its logistics operations, connecting fulfillment centers to last-mile delivery stations. We explores Amazon's capacity planning strategies for this crucial segment, its integration with demand forecasting, and its online fulfillment optimization routing. We highlight the challenges of managing a vast transportation network and the innovative solutions employed to enhance efficiency and reduce costs without sacrificing delivery speed.

3 - Causal Inference in Transportation Efficiency Estimation

Zhikun Gao, Amazon, Sammamish, WA, United States

Ground transportation plays a crucial role in Amazon's transportation network, for example using Line-Haul (LH) trucks. The modification of network connection and configuration will lead to changes in operation processes, which impacts transportation efficiency as a result. Learning the dependence relationships between operation processes and transportation efficiency helps us to quantify benefits of changing network designs, as well as identify opportunities to further improve operations. This talk introduce a generalized causal inference framework to attribute causal factors of transportation efficiency on operation processes. Through this framework, we can assess the impacts of network re-configurations and explore the potential for further improvement.

4 - Inventory Feasible and Capacity Compliant Facility Assignment Model

Arun Jotshi, Amazon, Seattle, WA, United States, Ayten Turkcan Upasani, Yufei Wang, Kaushik Krishnan

Amazon's fulfillment network is enormous in scale. For US alone there are hundreds of fulfillment warehouses (called fulfillment centers or FCs) located throughout the country where customer orders get fulfillment. The assignment of the orders placed by customers located across the country to the FCs is a critical decision as it generates the origin-destination (OD) flows of shipments which determine transportation cost and delivery speed. In this paper, we present FC Assignment Model (FCAM), a planning system that produces the optimized OD flows to guide the actual order assignment of the fulfillment execution systems. It is formulated as a linear programming problem with the objective is to minimize the total distance of OD flows subject to FC capacity and inventory availability constraints. The key challenge is the enormous product selection (hundreds of millions) which makes modeling individual products intractable. An important innovation of FCAM is the incorporation of inventory bounds at aggregated level without tracking individual products. Our back-testing with historical data shows that the optimized OD flows produced by FCAM are inventory feasible and capacity compliant with significant distance reductions compared with historical actuals without FCAM.

5 - Solving Large Scale Assignment Problems with Decomposition

Ozlem Bilginer, Amazon.com, SEATTLE, WA, United States

Various Amazon outbound execution systems need to allocate orders to paths subject to multiple constraints. The path decision for the orders includes fulfillment center choice, there are many paths out of each eligible fulfillment center, and paths are subject to capacity constraints. As such, these are large scale assignment problems, not easily solvable in their canonical form. In this talk, we present a decomposition algorithm to find primal and dual optimal solutions for one of these large scale linear assignment problems. We discuss the strategies for fast convergence, including problem reformulation and mixed use of various dual update methods.

TB61

Summit - Ballroom 3

Towards Pedagogical Excellence in the Era of Growing Diversity and Technological Mediation

Panel Session

Committee's Choice

Co-Chair: Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, United States

1 - Moderator Panelist

Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, PA, United States

2 - Panelist

Chrysafis Vogiatzis, University of Illinois Urbana-Champaign, Urbana, IL, United States

3 - Panelist

Ken Murphy, Merage School of Business, UC Irvine, Irvine, CA, United States

TB62

Summit - Signature Room

An Introduction to Decision Diagrams for Optimization

Invited Session

TutORial

Chair: Harish Krishnan, University of British Columbia, Vancouver, BC, Canada

1 - An Introduction to Decision Diagrams for Optimization

Willem-Jan van Hoeve, Carnegie Mellon University, Pittsburgh, PA, United States

This tutorial provides an introduction to the use of decision diagrams for solving discrete optimization problems. A decision diagram is a graphical representation of the solution space, representing decisions sequentially as paths from a root node to a target node. By merging isomorphic subgraphs (or equivalent subproblems), deci sion diagrams can compactly represent an exponential solution space. This ability can reduce solving time and memory requirements potentially by orders of magnitude. That said, exact decision diagrams can still be of exponential size for a given problem, which limits their practical applicability to relatively small instances. However, recent research has introduced a scalable approach by compiling polynomial-sized relaxed and restricted diagrams that yield dual and primal bounds, respectively. These can be combined in an exact search to produce a generic decision diagram-based branch and-bound method. This chapter describes how this approach provides a scalable solution method for state-based dynamic programming models. In addition, the chap ter shows how this approach can be applied to, and embedded in, other computational paradigms including constraint programming, integer programming, and column elim ination. After this chapter, readers will have an understanding of the basic principles of decision diagram-based optimization, an appreciation of how it compares it to other optimization methods, and an understanding of what types of optimization problems are most suitable for this new technology.

TB63

Regency - 601

Emerging Topics in IS and OM

Invited Session

Information Systems

Chair: Xingyu Fu, University of New South Wales, Sydney, 2113, Australia

Co-Chair: Pin Gao, The Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

1 - Fishing in the Same Pond: Campaign Interference in Digital Advertising

Francesco Balocco, Vrije Universiteit, Amsterdam, Netherlands, Yixin Lu, Bernd Skiera, Lennart Kraft

Display advertising has evolved into a sophisticated digital supply chain in the past decade. Advertisers increasingly rely on intermediaries known as Demand Side Platforms (DSPs) to streamline the complex tasks of running ad campaigns and purchasing digital ad spaces (impressions) via real-time auctions. DSPs automate the process of targeting and bidding, representing their clients in the digital ad marketplace.
Advertisers running concurrent campaigns for different marketing goals (e.g., brand awareness vs. retargeting) often employ multiple DSPs, each with its unique technology and expertise. While providing access to specialized tools for more effective campaign management, this approach may introduce inefficiencies.
Our research investigates how running campaigns via multiple DSPs undermines advertisers' revenues. Using a dataset of over 1.2 million ad auctions, we show that DSPs working for the same advertisers often inadvertently bid against each other for the same audience, driving up costs and leading to a decrease in the return on investment (ROI) for campaigns.
Based on this evidence, we formulate the optimal campaign combination selection problem in the presence of overlaps. Using actual instances from our ad auction dataset, we analyze the difference between the advertisers' realized revenues and their counterfactual revenue from using the optimal campaign combination. We finally perform computational experiments and approximate the advertisers' optimal campaign revenue using multi-armed bandits.

2 - Licensing and Contracting on the Intellectual Property Contests

Dayu Wu, University of Wisconsin-Madison, Madison, WI, United States

Innovation contests offer a novel and effective method for firms to source intellectual property (IP) from the public. A significant challenge for firms is designing an IP contest that both stimulates creativity and ensures profits. Firms must consider market demand and IP strategies, choosing between upfront and royalty-based licensing contracts to motivate participation and reduce sourcing costs. This study develops a novel contest model to determine optimal IP contest design, focusing on IP strategies and licensing contract decisions. We derive the optimal licensing mechanisms under various scenarios. Our findings indicate that the royalty licensing contract dominates upfront licensing contract as royalty rate increases. Innovators motivated by potential IP monetization exert extra effort to win the royalties licensing contract. However, with a low licensing rate, upfront licensing contracts boost more innovation effort than royalty-based contracts. Furthermore, we explore the effects of collaborative R&D between firms and innovators through IP contest. While reducing R&D costs, collaboration could have negative impacts on product quality, doing harm to the company profits. Our analysis provides insights into collaborative innovation ecosystems and effective IP management, highlighting the balance between fostering innovation and maintaining product standards.

3 - Monetary vs Attention Subsidies on Two-Sided Platforms

Chang Dong, Durham University, Durham, United Kingdom

Facilitating the creation of high-quality content is critical to the success of two-sided platforms. We develop an analytical framework about the equilibrium consequences of two canonical types of subsidy policies in platform governance: providing additional monetary rewards or audience attention to content creators. Cultivating and identifying quality content providers is time-consuming, and these content providers often have limited lifecycles. With a dynamic model, we examine both the short-term and the long-term impacts of the two subsidy policies. Recognizing the intergenerational linkage between new and old content providers has a profound impact on the market equilibrium and the effect of platform governance policies. Our work provides concrete guidelines to help platform managers design subsidy policies to encourage content creation.

TB64

Regency - 602

Data Analysis and Optimization in Communication and Social Networks

Invited Session

Social Media Analytics

Chair: Alexander Semenov, University of Florida, Shalimar, FL, United States

Co-Chair: Nehir Tanyel, University of Texas at San Antonio, San Antonio, TX, 78249, United States

Co-Chair: Chinju Paul, Mississippi State University, Starkville, MS, 39762, United States

1 - An Optimization Framework for Constructing Diversified Portfolios in Market Graphs

Maciej Rysz, Miami University, Oxford, OH, United States

Integration of market graphs with traditional portfolio optimization models is a relatively unexplored area. Existing studies predominantly focus on either portfolio optimization or the construction of market graphs in isolation, with limited exploration of the underlying properties when combining both. This work considers a general risk-reward portfolio optimization model that captures interactions between different assets using a market graph. Focus is put on the interplay between risk measures, graph types, optimal set structures, and portfolio performance. In particular, the model is explored within the context of higher-moment coherent risk measures, correlation and sign market graphs, with the goal of identifying diversified sets of assets that form distance-based independent sets in the underlying graphs. Computational results demonstrate that integrating a graph structure can significantly improve the performance over standard portfolio optimization models and help avoid overfitting. Overall, this study contributes to a deeper understanding of the potential benefits derived from utilizing graph structures in stochastic optimization models.

2 - Global Optimization Approaches to Solving Large Nonlinear Equation Systems: Applications and Techniques

Alexander Semenov, University of Florida, Shalimar, FL, United States, William Trevena, Michael Hirsch, Panayote (Panos) Pardalos

Solving systems of nonlinear equations, even small ones, can be quite challenging. As the size of the systems increases, the complexity of finding all solutions can grow dramatically. This research explores the transformation of such systems into global optimization problems and utilizes efficient algorithms to efficiently find all solutions. The talk includes examples demonstrating the approach on large systems. We experiment with derivative-free optimization algorithms alongside gradient-based algorithms, utilizing automatic differentiation.

3 - The Polarizing Effect of News Playlists on YouTube

Nehir Tanyel, University of Texas at San Antonio, San Antonio, TX, United States

This study delves into the polarization effects observed within news playlists established by prominent channels like Fox and CNN, focusing on contentious topics such as war. Unlike personalized playlists, these curated collections offer a unique lens into the dissemination of polarizing viewpoints by influential media outlets. Through a comprehensive analysis of content composition, viewer engagement, and the amplification of divergent narratives, this research aims to uncover the mechanisms driving polarization within these curated news playlists on YouTube.

4 - Social Media Use and Mental Well-being: The Effect of Motives and Use Behavior

Chinju Paul, Mississippi State University, Starkville, MS, United States

This study delves into the impact of social media usage on depression and mental well-being. While existing literature suggests a link between social media use and heightened depression alongside decreased well-being, some studies propose a positive correlation with well-being. The inconclusive finding warrants an examination of how individuals engage with social media. The literature identifies two main types of use – active use and passive use. Active use involves interacting with content by posting, commenting, or engaging in political causes. Conversely, passive use involves simply scrolling through content without interaction. Our preliminary analysis reveals that active social media use reduces depression levels, whereas passive use exacerbates them. Moreover, different types of active use and passive use result in varied effects. To deepen our understanding, we explore users' motives for engaging with social media and how they align with their usage patterns. We hypothesize that when users' motives align with their behavior, it enhances well-being, irrespective of active or passive engagement. Conversely, misalignment leads to poorer mental well-being and may result in excessive social media usage due to reduced satisfaction. To empirically test these hypotheses, we will conduct surveys and interviews. This research offers both theoretical and practical contributions. By providing comprehensive insights into how social media usage patterns influence mental well-being, it enriches existing literature in the field. Additionally, this study can offer users insights to modify their social media use patterns to better align with their motives and derive benefits from their social media use.

5 - Insecure? How Ephemerality Affects Privacy and Self-Misrepresentation on Online Social Networks: A Moderated Moderation Model

Siwei Jiang, Mississippi State University, Starkville, MS, United States

Online social networks (OSNs) provide users with great opportunities to interact with others while selectively disclosing themselves by building a digital self-image. In privacy literature, privacy concerns have been demonstrated as one of the reasons that cause users to provide fabricated information on OSNs. To reduce users' privacy concerns and facilitate more authentic self-representations, some OSNs introduced an ephemerality feature, which limits the visibility of users' information or content only for a short timeframe. However, psychology scholars show that insecure users especially young teenagers/adults are more prone to use impression management to achieve a desirable self-image on OSNs not only for social connections or support but also for potential social bonds and perceived traits based on their self-deceptive communications. Drawing upon the privacy calculus and self-impression perspective, this study examines the three-way interaction between privacy concerns and the ephemerality feature with users' insecurity states. We hypothesize that the ephemerality feature can effectively reduce users' privacy concerns and then lead to more willingness to disclosure intention, but this moderating effect varies contingent on the types of users' security states. We plan to conduct a survey study to empirically validate the hypotheses. Our study is expected to have both theoretical and practical implications. We contribute to the privacy literature by explaining users' self-disclosure behavior from an insecure emotional perspective and ephemerality of the content. For practitioners, this study indicates that OSNs should reconsider their content-sharing strategies by considering the mental well-being of users to foster a more authentic and transparent communication environment.

TB65

Regency - 603

K-12 Outreach Panel: Collaboration Across Institutions

Panel Session

Education Outreach

Co-Chair: Fenglian Pan, University of Arizona, Tucson, AZ, United States

Co-Chair: Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States

1 - Moderator Panelist

Zihan Zhang, ISyE Georgia Tech, Atlanta, GA, United States, Kara Combs, Jian Liu, Ana Muriel, Tuba Ketenci, Fenglian Pan

2 - Moderator Panelist

Fenglian Pan, University of Arizona, Tucson, AZ, United States

3 - Panelist

4 - Panelist

Ana Muriel, University of Massachusetts, Amherst, MA, United States

5 - Panelist

Jian Liu, University of Arizona, Tucson, AZ, United States

6 - Panelist

Kara Combs, Air Force Research Laboratory, Englewood, OH, United States

TB66

Regency - 604

ENRE Awards Session

Award Session

Energy, Natural Resources, and the Environment Special

Chair: Sandra Eksioglu, University of Arkansas, Fayetteville, AR, United States

TB67

Regency - 605

Innovations in AI for Optimization and Control

Contributed Session

Chair: Yujie Wang, University of Miami, 8255 SW 72nd Ct, Miami, FL, 33143, United States

1 - Interactive Real-time Decision Support in Digital Control Rooms: An Empirical Study on Generative AI Implementation

Marijn Verschelde, IÉSEG School of Management, Lille, France, Kristof Coussement, Léon Sobrie

In this paper, we explore the use of generative AI to provide customizable, explainable real-time decision support for operational end users who must make split-second decisions under variable time constraints. Our empirical study, conducted within the context of digital railway control room management at Infrabel, addresses the challenges posed by a constant stream of tasks with both expected and unexpected peaks. Traditional visualizations of local and global SHAP (Shapley Additive Explanations) have been found to be too time-consuming to interpret and insufficiently tailored to the specific needs of operators and settings. We evaluate the effectiveness of recent advancements in generative AI in enhancing XAI-based predictive analytics, enabling setting- and operator-specific customization and interactive decision support.

2 - Enterprise-Wide Optimization Using AI-Driven Mathematical Programming

Carlos Paternina-Arboleda, San Diego State University, San Diego, CA, United States, Jesus Velasquez

Mathematical programming (MP) is a cornerstone of optimization across various disciplines. However, formulating effective MP models often requires significant human expertise. This article explores the intersection of generative artificial intelligence (AI) and MP, proposing a novel framework termed Generative Mathematical Programming Artificial Intelligence (GMP-AI). GMP-AI leverages AI's ability to learn patterns from data to automate MP model generation. The framework empowers the AI to learn the relationships between problem data and the corresponding model structure. This paves the way for GMP-AI to generate new MP models for unseen optimization problems within the learned domain. The article discusses the potential applications of GMP-AI in various fields and concludes by outlining promising avenues for future research in this emerging area. Although our work has made great advancement in automating the process of building large-scale mathematical models, based on the libraries created, there is still a lot of potential for growth in this promising field.

3 - Asymptotic Behavior of Adversarial Training Estimator Under L_\Infty-Perturbation

Yiling Xie, Georgia Tech, ATLANTA, GA, United States, Xiaoming Huo

Adversarial training has been proposed to hedge against adversarial attacks in machine learning and statistical models. This paper focuses on adversarial training under ℓ_∞-perturbation, which has recently attracted much research attention. The asymptotic behavior of the adversarial training estimator is investigated in the generalized linear model. The results imply that the limiting distribution of the adversarial training estimator under ℓ_∞-perturbation could put a positive probability mass at 0 when the true parameter is 0, providing a theoretical guarantee of the associated sparsity-recovery ability. Alternatively, a two-step procedure is proposed---adaptive adversarial training, which could further improve the performance of adversarial training under ℓ_∞ perturbation. Specifically, the proposed procedure could achieve asymptotic unbiasedness and variable-selection consistency. Numerical experiments are conducted to show the sparsity-recovery ability of adversarial training under ℓ_∞-perturbation and to compare the empirical performance between classic adversarial training and adaptive adversarial training.

4 - A Novel Heterogeneous Recurrence Network Analysis for Spatiotemporal Data

Yujie Wang, University of Miami, Coral Gables, FL, United States, Cheng-Bang Chen

Recent advancements in recurrence analysis have significantly enhanced our understanding of dynamical systems, uncovering complex, non-linear, and non-stationary properties in various fields such as manufacturing, healthcare, and social sciences. Despite these advances, there is a notable gap in applying recurrence methods to spatiotemporal data, which is prevalent in nature and contains diverse patterns worth exploring. To address this, we propose a novel methodology: heterogeneous recurrence network analysis for spatiotemporal data. This approach processes data chronologically while preserving spatial information, thereby maximizing data utility. We validate our methodology by analyzing the high dimensional EEG signals to distinguish human emotions over time. This innovative application not only aids in mental health diagnosis but also provides deeper insights into spatiotemporal dynamics.

TB68

Regency - 606

Strategies in ESG and Contractual Decisions

Contributed Session

Chair: WENTING LI, Arizona State University, Tempe, AZ, 85044-1746, United States

1 - Optimal Contract Selection Strategies Under Three Different Structures

Qing Zhu, Nanjing University of Science and Technology, nanjing, China, People's Republic of, Yudong Wang, Stephen Disney, Wei Xing

We explore optimal contract choice under monopoly, retailer competition, and manufacturer competition structures. The impact of interest rates and the substitution rate between two products on manufacturers' and retailers' profits is also considered. We also focus on the similarities and differences between the manufacturers and the retailers under Bertrand and Cournot competition. We identify the contracts the manufacturers wish to adopt, the contracts the retailers prefer manufacturers to choose, and whether there are differences in the reasons for contract selection. Our model employs a two-stage Stackelberg game. The research findings indicate that in the monopoly case, with retailers engaging in either Bertrand and Cournot competition, the volatility of spot market price affects the manufacturer's contract choice and the retailers' contract preferences in the same manner. We found that under manufacturer competition, whether in Bertrand or Cournot games, the manufacturer's contract choices and equilibrium strategies are the same. However, in this structure, the retailers' preferences for the type of contract chosen by manufacturers may vary depending on the magnitude of market demand price volatility. Furthermore, the study concludes that under the two game modes of manufacturer and retailer competition, both manufacturer and retailer profits decrease with an increase in the substitutability between the two products, and manufacturer profits decrease with interest rate increases.

2 - ESG Data Imputation and Greenwashing

Giulia Crippa, Princeton University, Princeton, NJ, United States

In recent years, there has been a notable surge of Environmental, Social, and Governance (ESG) investing. This paper provides a simple and comprehensive tool to tackle the issue of missing ESG data. Firstly, it allows to shed light on the failure of ESG ratings due to data sparsity. Exploiting machine learning techniques, we find that the most significant metrics are promises, targets and incentives, rather than realized variables. Then, data incompleteness is addressed, which affects about 50% of the overall dataset. Via a new methodology, imputation accuracy is improved with respect to traditional median-driven techniques. Lastly, exploiting the newly imputed data, a quantitative dimension of greenwashing is introduced. We show that when rating agencies do not efficiently impute missing metrics, ESG scores carry a quantitative bias that should be accounted by market players.

3 - Relative vs. Absolute Volume-Based Export Restrictions: Choice, Design, and Effectiveness

Lucy Chen, National University of Singapore, Singapore, Singapore, Srinagesh Gavirneni, Aditya Vedantam, Shuguang Zhang

We study a firm’s production and export decisions when facing export restrictions imposed by the government, who aims to maximize the social welfare. We investigate how different export restrictions affect the firm’s decisions and profitability as well as the social welfare.

4 - A Procurement Advantage in Disruptive Times: New Perspectives on ESG Strategy and Firm Performance

Wenting Li, Arizona State University, Tempe, AZ, United States, Yimin Wang

This study investigates whether elevated environmental, social, and governance (ESG) practices in the global supply chain confer firm resilience during disruptions, and, if so, the underlying mechanisms. Drawing on the COVID-19 pandemic as a natural experiment, we define a firm's resilience as its relatively superior financial performance during the pandemic. To isolate the causal effect of ESG practices on a firm's resilience, we employ an instrumental variable estimation approach, exploiting exogenous variation in ESG practices driven by firms' exposure to political orientations. The results reveal that increased ESG practices strengthen a firm's resilience during disruptions: a 1% increase in ESG practice scores leads to a 0.215% increase in firms' return on assets. We analyze the mechanisms driving this resilience effect and show that improved ESG practices are associated with reduced purchasing costs and higher profitability amid disruptions. Through a series of mechanism tests exploiting heterogeneity of stakeholder resources, we provide robust evidence that ESG enhances operational congruency with suppliers, which becomes critical in securing a procurement advantage during severe external constraints. Contrary to popular belief, there is little evidence that the ESG improves price premiums during the disruption. Therefore, strengthening operational alignment with suppliers is more critical than fostering customer loyalty with ESG labels, highlighting the primacy of ESG-enabled upstream stakeholder collaborations. Managerially, this research quantifies the economic rationale for ESG practices, offering managers empirical evidence on the imperative to prioritize specific stakeholder alignments as opposed to broad, vague ESG practices that may harm firm performance.

TB69

Regency - 607

SpORts II: SpORts Betting and Social Impact

Invited Session

spORts

Chair: Jay Simon, American University, Washington, DC, United States

1 - Inefficiencies In the Betting Market for Popular Events? Evidence from the Ncaa Men&Rsquo;S Basketball Tournament

Dan Hickman, University of Idaho, Moscow, ID, United States

The NCAA men’s basketball tournament, commonly referred to as March Madness, generates a significant amount of betting activity. The rise in legal access to sports gambling in the United States, along with the popularity of the event mean that a substantial number of those taking part in the market may not be regular market participants. Is it possible for an informed participant to exploit potential biases in the market? This study analyzes outcomes from the past 30 years to determine whether factors such as regular season performance, travel distances, game time, recency bias or others lead to inefficiencies in the market.

2 - Betting Line Movement and Inefficiencies in Four Major spORts

Jay Simon, American University, Washington, DC, United States

This work explores real-time betting line movement in four major sports. Particular attention is given to how both new and previously observed inefficiencies evolve over the duration of the betting window for each game. While sportsbooks' prices are generally reliable, they are not perfect, and several inefficiencies arise.

3 - Online spORts Gambling and Individual Wellbeing

Mike Gordon, Virginia Tech, Blacksburg, VA, United States

While legalization has led to an explosion in the popularity of online sports gambling, research has not evaluated the impact of the change in access to sports gambling on societal welfare. We evaluate the impact of the legalization and digitization of online gambling on mental health, financial, and crime outcomes. We look to understand how nascent avenues for gambling affect consumers at large. We additionally look to address potential disparities in the population and how they affect this behavior.

4 - Carbon Footprint Impact of NCAA Conference Realignment

Sia Sheguri, Arizona State University, Tempe, AZ, United States, Jiayi Hong, Ross Maciejewski, Ron Askin

With the recent realignments in the National Collegiate Athletic Association's Power Five conferences, there has been a surge in travel distances between games, thereby elevating the carbon footprint dramatically. In this work, we investigate the impact of conference realignment on carbon footprint using team travel for football games as a proportional surrogate for impact. Using the geographic location of each team and typical travel method of air for trips exceeding 200 miles (bus for shorter trips), along with historical and proposed schedules, the carbon footprint produced by Power Five team schedules is computed before and after recent realignments. A model is developed to optimally divide each conference into divisions and the potential savings of using divisions as opposed to random within conference scheduling is shown. We believe that these strategic sub-divisions could be a step towards a greener collegiate sports community and mitigate the impact of realignment. Finally, for comparison, a model is developed to allocate the set of Power Five teams into an optimal set of conferences for minimizing carbon footprint. The model can be extended to include factors such as desired population coverage or other economic constraints to ensure recruiting and exposure goals.

TB70

Regency - 701

Optimization and Machine Learning in Power Systems

Invited Session

ENRE: Electricity

Chair: Minseok Ryu, Arizona State University, Tempe, AZ, United States

1 - Physics-Informed Heterogeneous Graph Neural Networks for DC Blocker Placement

Hongwei Jin, Argonne National Laboratory, Lemont, IL, United States

The threat of geomagnetic disturbances (GMDs) to the reliable operation of the bulk energy system has spurred the development of effective strategies for mitigating their impacts. One such approach involves placing transformer neutral blocking devices, which interrupt the path of geomagnetically induced currents (GICs) to limit their impact. The high cost of these devices and the sparsity of transformers that experience high GICs during GMD events, however, calls for a sparse placement strategy that involves high computational cost. To address this challenge, we developed a physics-informed heterogeneous graph neural network (PIHGNN) for solving the graph-based dc-blocker placement problem. Our approach combines a heterogeneous graph neural network (HGNN) with a physics-informed neural network (PINN) to capture the diverse types of nodes and edges in ac/dc networks and incorporates the physical laws of the power grid. We train the PIHGNN model using a surrogate power flow model and validate it using case studies. Results demonstrate that PIHGNN can effectively and efficiently support the deployment of GIC dc-current blockers, ensuring the continued supply of electricity to meet societal demands. Our approach has the potential to contribute to the development of more reliable and resilient power grids capable of withstanding the growing threat that GMDs pose.

2 - A Bidding Strategy Algorithm for Distributed Energy Resource (DERs) Aggregators: Participation in Day-ahead and Real-time Energy and Ancillary Markets

Sara Ostovar, Arizona State University, Tempe, AZ, United States, Mojdeh Hedman

Federal Energy Regulatory Commission (FERC) Order No. 2222 paved the way for Distributed Energy Resource (DERs) to participate in the wholesale market via aggregations to meet regulatory requirements. However, it is necessary to develop bidding strategies unique to DER aggregators due to the uncertainty and low operational cost. Multiple work in literature investigate the bidding strategy of DER Aggregators, considering the risk of failing obligations which pose financial risks to the aggregator.
This work proposes a bidding strategy for a DER aggregator whose resource mix comprises Photovoltaic (PV) units and Energy Storage Resources (ESRs) participating in the day ahead and real-time energy and ancillary service market. This work covers the day-ahead model with a vision of real-time as well as a separate real-time bid generation module. Regulation-up is considered concerning the participation of DERs in provision of ancillary services. The framework uses a scenario-based modeling approach for market price and PV output power uncertainty. The model produces price-quantity offer curves for the day-ahead and real-time energy and ancillary service market.

3 - Grid-Aware on-Route Fast-Charging Infrastructure Planning for Battery Electric Bus with Equity Considerations: a Case Study in South King County

Xinyi Zhao, University of Washington, Seattle, WA, United States, Chaoyue Zhao, Grace Jia

The transition from traditional bus fleets to zero-emission ones necessitates the development of effective planning models for battery electric bus (BEB) charging infrastructure. On-route fast charging stations, distinct from on-base charging stations, present unique challenges related to safe operation and power supply capacity, making it difficult to control grid operational costs. This paper establishes a novel framework that integrates the bus route network and power network, which leverages the inter-dependency between both networks to optimize the planning outcomes of on-route BEB charging stations in South King County. The problem is formulated as a mixed-integer second-order cone programming model, aiming to minimize the overall planning cost, which includes investments in charging equipment, power facility, and grid operation. Furthermore, fairness measurements are incorporated into the planning process, allowing for the consideration of both horizontal transit equity and vertical transit equity based on different zone merging criteria within the county's existing census tracts. The results of this planning model offer valuable insights into achieving both economic efficiency and social justice in the design of on-route charging facilities for BEBs in South King County.

4 - Scalable Algorithms for Mitigating Impacts of Geomagnetic Disturbances on Electric Grids

Joshua Clugston, Arizona State University, Tempe, AZ, United States, Minseok Ryu

Space weather introduces geomagnetically induced currents (GICs) into the electrical grid, potentially disturbing grid operation and causing damage to grid infrastructure. The risks imposed by GICs can be reduced by the appropriate placement of GIC blocking devices. However, optimal placement of such devices typically leads to solving a challenging mixed-integer nonlinear program (MINLP). To address the computational challenges posed, we therefore propose an alternating direction method of multipliers (ADMM) based heuristic algorithm, developed by exploiting the structure of the blocker placement problem. We demonstrate on the ACTIVSg2000 synthetic grid test case that our heuristic algorithm produces high-quality solutions for moderately sized blocker placement problems without expending substantial computational resources.

5 - Electricity price forecasting in a power network using GNN-based encoder-decoder architecture

Krishna Rijal, University of South Florida, Tampa, FL, United States, Nicolas Bustos, Tapas Das

Electricity price forecasting (EPF) is a critical function in the electricity market due to its potential impact on resource planning and allocation, cost management, and profit maximization. The inherent complexity arising from fluctuating supply from diverse sources and variable demand patterns presents a challenge in accurate electricity price prediction. In this study, we propose an approach leveraging graph neural networks (GNNs) for EPF integrating factors such as supply and demand variations, network power flow, and other exogenous variables. The graph representation considers the demand, generation, weather information, and fuel costs as the node attributes while the power flow is considered as the edge attribute. The model architecture also considers a self-supervised learning mechanism designed to detect price spikes, which is likely to enhance the accuracy of our price forecasts.

TB71

Regency - 702

Data Assimilation for Simulation Optimization

Invited Session

Data Mining

Chair: Runsang Liu, The Pennsylvania State University, University Park, PA, United States

Co-Chair: Hui Yang, Pennsylvania State University, University Park, PA, United States

1 - Recurrence Network Modeling for Cardiac Digital Twin

Runsang Liu, The Pennsylvania State University, University Park, PA, United States

The paradigm shift from population-based decision-making towards precision medicine has led to a plethora of imaging data of the heart. While most of these accumulated data are underutilized, there is a need to develop effective representations of heart geometry and further enable the integration of different types of clinical data to dissect disease heterogeneity and assist intervention planning. In this paper, we present a new recurrence network representation of the heart. Each point in the heart geometry is treated as a vertex, and the edges are formed based on the spatial proximity of vertices. By simulating the network as a physical system, the network self-organizes into a steady topology by minimizing system energy. Experimental results show the effectiveness of the recurrence network model of heart by reconstructing both the external and internal structures of the heart.

2 - Multi-Source Data and Knowledge Fusion via Deep Learning for Dynamical Systems: Applications to Spatiotemporal Cardiac Modeling

Bing Yao, The University of Tennessee Knoxville, Knoxville, TN, United States

Advanced sensing and imaging provide unprecedented opportunities to collect data from diverse sources for increasing information visibility in spatiotemporal dynamical systems. Furthermore, the fundamental physics of the dynamical system is commonly elucidated through a set of partial differential equations (PDEs), which plays a critical role in delineating the manner in which the sensing signals can be modeled. Reliable predictive modeling of such spatiotemporal dynamical systems calls upon the effective fusion of fundamental physics knowledge and multi-source sensing data. This paper proposes a multi-source data and knowledge fusion framework via deep learning for dynamical systems with applications to spatiotemporal cardiac modeling. This framework not only achieves effective data fusion through capturing the physics-based information flow between different domains, but also incorporates the geometric information of a 3D system through a graph Laplacian for robust spatiotemporal predictive modeling. We implement the proposed framework to model cardiac electrodynamics under both healthy and diseased heart conditions. Numerical experiments demonstrate the superior performance of our framework compared with traditional approaches that lack the capability for effective data fusion or geometric information incorporation.

3 - Small Data and Sparse Bayesian Learning for Sequential Inference of Network Connectivity

Jinming Wan, Binghamton University, Vestal, NY, United States

While significant efforts have been attempted in the design, control, and optimization of complex networks, most existing works assume the network structure is known or readily available. However, the network topology can be radically recast after an adversarial attack and may remain unknown for subsequent analysis. In this work, we propose a novel Bayesian sequential learning approach to reconstruct network connectivity adaptively: A sparse Spike and Slab prior is placed on connectivity for all edges, and the connectivity learned from reconstructed nodes will be used to select the next node and update the prior knowledge. Central to our approach is that most realistic networks are sparse, in that the connectivity degree of each node is much smaller compared to the number of nodes in the network. Sequential selection of the most informative nodes is realized via the between-node expected improvement. We corroborate this sequential Bayesian approach in connectivity recovery for the IEEE-118 power grid system. Results indicate that only a fraction (~50%) of the nodes need to be interrogated to reveal the network topology.

4 - Digital Twinning and Optimization of Manufacturing Process Flows

Hankang Lee, The Pennsylvania State University, University Park, PA, United States, Hui Yang

The new wave of Industry 4.0 is transforming manufacturing into data-rich environments. This provides an unprecedented opportunity to feed large amounts of sensing data collected from the physical factory into the construction of digital twin (DT) in cyber-space. Despite this potential, there has been limited exploration of how DT technology can enhance the efficiency and autonomy of small and medium-sized manufacturing enterprises. Therefore, there is an urgent need to exploit the full potential of data analytics and simulation-enabled DTs for advanced manufacturing. Hence, this study presents the design and development of DT models for simulation optimization of manufacturing process flows. First, we develop a multi-agent simulation model describing nonlinear and stochastic dynamics among a network of interactive manufacturing things, including machines, automated guided vehicles (AGVs), queues, and jobs. Second, we propose a statistical metamodeling approach to design sequential computer experiments aimed at optimizing AGV utilization under uncertainty. Third, we construct two new graph models—job flow graph and AGV traveling graph—to monitor and analyze the real-time performance of manufacturing jobshops. The effectiveness of the proposed simulation-enabled DT approach is assessed through experimental studies. Results show that the proposed methodology effectively transforms conventional manufacturing jobshops into novel DT-enabled smart factories. The sequential design of experiments significantly reduces the computational overhead of expensive simulations while efficiently scheduling AGVs to achieve cost-effective production throughput. This research is strongly promised to help manufacturers fully utilize big data and DT technology to gain competitive advantages in the global marketplace.

TB72

Regency - 703

Advancements in Medical Data Analysis and Intervention

Contributed Session

Chair: Gunsu Dagistanli, N/A

1 - Enhanced Laryngeal Cancer Screening Using Deep Learning: Integrating Phoneme Analysis, Diverse Sentence Structures, and Patient Data

Yong Oh Lee, Hongik University, Seoul, Korea, Republic of

Symptoms of laryngeal cancer manifest primarily through voice changes, and diagnosis typically relies on laryngoscopy. This process lacks objective indicators for voice changes, making early detection difficult. Recent advances in deep learning have opened up new possibilities for voice-based laryngeal cancer screening, but its clinical application is still limited due to its relatively low accuracy.
In this presentation, we propose a multi-faceted approach to address these issues and improve screening performance. First, we adopt a phoneme-based detection approach using the /a:/ phoneme and compare two feature extraction methods: Mel-frequency correlation coefficient (MFCC) and octave-frequency correlation coefficient (OFCC). The results show that OFCC achieves higher performance than MFCC in laryngeal cancer detection. Second, we extend our analysis to include different sentence forms without relying solely on the /a:/ phoneme. This allows us to capture a wider range of vocalization characteristics affected by laryngeal cancer, providing a more

comprehensive voice screening. Third, we integrate patient information such as age, gender, and medical history with the results of voice signal analysis using convolutional neural network (CNN) models. The fusion of these data significantly improves the accuracy of the screening tool.
Experimental results show that this integrated approach can accurately distinguish laryngeal cancer more than 90% of the time.

2 - Binary Gaussian Copula Synthesis: Pioneering Data Augmentation for Early Dialysis Intervention in CKD Patients

Hamed Khosravi, West Virginia University, Morgantown, WV, United States, Srinjoy Das, Mohammad Abdullah Al-Mamun, Imtiaz Ahmed

The Center for Disease Control estimates that over 37 million US adults have chronic kidney disease (CKD), with 90% unaware of their condition due to symptomless early stages. This disease severely affects life quality, particularly as it progresses to requiring dialysis. Early dialysis prediction is crucial for improving patient outcomes and aiding healthcare providers in timely decision-making. However, developing an effective machine learning (ML)-based Clinical Decision Support System (CDSS) for such early prediction is challenging due to data imbalance—there are significantly fewer dialysis than non-dialysis CKD patients, which hampers the accuracy of ML models.

To tackle this, we introduce a novel approach named Binary Gaussian Copula Synthesis (BGCS), specifically designed for binary medical datasets. BGCS excels at generating synthetic data that preserves the original distribution, enhancing the detection of potential dialysis patients. Our results show that BGCS outperforms traditional methods, improving dialysis prediction across various ML models. For example, using a Random Forest model, BGCS achieved a 72% improvement over state-of-the-art augmentation methods, demonstrating significant enhancements even compared to situations without augmentation. Additionally, we developed an ML-based CDSS using decision tree models, favored for their interpretability in healthcare. This system is engineered to assist clinicians in proactive decision-making and effective treatment planning for CKD patients likely needing dialysis soon. Through detailed feature analysis and careful data preparation, our CDSS delivers accurate, actionable predictions, serving as a vital tool in CKD management and treatment.

3 - Matching mechanism and queuing model in organ transplantation

Nani Zhou, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

Organ transplantation is the most effective treatment for many end-stage cancers. In the case of kidney transplantation, for example, most patients with end-stage renal disease wait in line on the deceased donor waiting list. ABO blood group compatibility is one of the main indicators of the match between a cadaveric kidney and a patient. The O blood group is the universal donor and the AB blood group is the universal recipient. Since the demand for kidneys is much higher than the supply, how to equitably distribute the limited kidney resources to individual patients is an urgent issue that needs to be improved. In the United States, for example, the proportion of blood groups O and A in the population is about 45% and 40%, and the corresponding proportions of blood groups in patients with end-stage renal disease are about 54% and 27%. As O kidneys are available to A patients, the probability of O patients receiving a kidney decreases, the average waiting time increases, and the dropout rate increases (patient deaths). The difficulty of the problem lies in how to measure the fairness of the queuing system and how to carve out the entire stochastic system. When the changes in the state of the system are sliced finely enough by time, we approximate this queuing process as a flow model and explore how to construct fairer kidney matching rules.

4 - Chemotherapy Appointment Scheduling at Hacettepe Oncology Hospital

Gunsu Dagistanli, Middle East Technical University, Ankara, Turkey, Meral Azizoglu

As the demand for chemotherapy treatment rises, efficient planning becomes critical for oncology clinics and hospitals. The first stage of planning, involves allocating patients to treatment days within the constraints of their treatment plans and hospital schedules. The second stage involves preparing daily patient and nurse schedules while adhering to chair limitations and maximizing patient satisfaction.

In this study we address the second stage problem, known as Chemotherapy Appointment Scheduling (CAS) problem, based on our observations at Hacettepe Oncology Hospital in Ankara. We present a deterministic Mixed-Integer Linear Programming (MILP) model with continuous appointment times for various defined single criterion or multi criteria objectives. The results of our computational experiment reveal that our model can handle up to problem instances with 10 patients, 3 nurses and 3 chairs. We also extend the problem to include infusion duration related stochasticity and propose a two-stage Stochastic Mixed Integer Programming (SMIP) model for its solution.

TB73

Regency - 704

Funding Opportunities on Data Science Research

Panel Session

Data Mining

Co-Chair: Ying Lin, University of Houston, Houston, TX

Co-Chair: Na Zou, University of Houston, Houston, TX, United States

1 - Moderator Panelist

Na Zou, University of Houston, Houston, TX, United States

2 - Moderator Panelist

Ying Lin, University of Houston, Houston, TX, United States

3 - Panelist

Cornelia Caragea, National Science Foundation, Alexandria, VA, United States

TB74

Regency - 705

Market Design, Analysis, and Optimziation of Zero-carbon Electricity Market

Invited Session

ENRE: Energy-Climate

Chair: Dongwei Zhao, Argonne National Laboratory, Westmont, United States

1 - Assessing Austria's 2040 Energy Storage Requirements

Robert Gaugl, Institute of Electricity Economics and Energy Innovation/Graz University of Technology, Graz, Austria, Sonja Wogrin

As Austrian decarbonization initiatives emphasize enhanced electrification, there is a notable lack of quantitative estimates/goals regarding necessary storage capacities. In this talk, we examine the expected need for short-, medium-, and long-term energy storage applications in the Austrian power system of 2040. We use a European Net-Transfer-Capacity model, featuring a single node per country, while adopting a finer spatial resolution for Austria. Multiple sensitivity analyses explore the impacts of different weather years, import/export capacities and costs. The study's findings underscore two main points: the pressing need for expanded storage capacity and the significant impact of allowing renewable energy curtailment on storage requirements.

2 - Evaluating Robust Technology Investment Portfolios to Achieve Zero-Carbon Power Systems

Shantanu Chakraborty, Massachusetts Institute of Technology, Cambridge, MA, United States, Audun Botterud, Jessika Trancik

The transition to low-carbon power systems necessitates substantial integration of weather-driven energy resources, such as wind and solar power. To ensure reliable supply of power over time and space, a portfolio of supply-side options and transmission infrastructure is required. While multiple technologies can be considered to provide these capabilities, there is a lack of comprehensive investigation into their trade-offs and complementarities, particularly under parametric uncertainties such as cost assumptions. This research aims to address this gap, offering insights to help identify robust, high-performing technology portfolios that can support future zero-carbon power systems.

This study employs a comparative approach using the Zero-emissions Electricity system Planning with HourlY operational Resolution (ZEPHYR), Sustainable Energy Systems (SES), and GenX models to analyze the roles of energy storage, transmission, demand response, and power generation at national, regional, and local scales. Additionally, we explore parametric uncertainties related to technology cost evolution and potential constraints on transmission infrastructure expansion. We investigate how the optimal deployment of power supply-side options depends on resource availability factors, the ability to expand transmission infrastructure, and the power and energy capacity cost evolution of energy storage. By synthesizing the differences across models, the study identifies favorable technology portfolios while accounting for uncertainties to provide a comprehensive framework that is aimed at guiding investments and policy decisions towards a zero-carbon power system for the US.

3 - Incentivizing long-term low-carbon power system resource adequacy under uncertainty: Accouting for the impact of weather information and extreme events

Shen Wang, Massachusetts Institute of Technology, Catonsville, MD, United States

To address global climate impacts, the electric power system is facing the challenges of achieving grid decarbonization while ensuring reliable power supplies. The transition toward clean renewable technologies, notably wind and solar which are characterized by uncertainty of power generation, could raise electricity supply reliability issues as they may not reliably dispatch electricity when needed. Furthermore, it is also argued that just penalizing outages of large fossil fuel power plants by conventional capacity derating is insufficient to fully account for the risks they introduce, as these plants may experience lengthier periods of outages and exhibit correlated vulnerabilities to extreme weather (i.e., winter blackouts in Texas and heatwave blackouts in California). The capacity accreditation mechanisms that lie at the heart of many present designs of capacity markets may not effectively promote efficient long-term investments, due to weak linkages of planning elements and operational uncertainties driven by weather information. Therefore, it is essential to improve the traditional deterministic planning and resource accreditation methods to capture the impacts of evolving weather-driven extremes on power systems. In this paper, we aim to address these issues through the following research questions below: 1) How can we incorporate weather-informed uncertainty in generation expansion planning, utilizing techniques such as stochastic or robust optimization? 2) Does considering uncertainty in renewable availability and correlated thermal power shutdowns significantly enhance the reliability and cost-effectiveness of resource plans? 3) What capacity credit definitions, informed by weather data, can accurately assess resource adequacy and be the basis of efficient incentives?

4 - 2032 miniWECC: Market Models for Zero Marginal Cost Resources

Allison Campbell, Pacific Northwest National Laboratory, Richland, WA, United States

With increased penetration of inverter-based renewable generation, next-generation power market models will need to evolve to account for the physics and economics of these units. This will require incorporating two new components to facilitate seamless commitment and dispatch of variable renewable generation while supporting system stability: responsiveness to power electronics control setpoints, and adaptation to zero marginal fuel costs. Development of new market model formulations that are responsive to these two conditions requires a testbed system that is both representative of real-world topologies and lightweight enough to facilitate rapid prototyping, but current market model test cases are either purely synthetic or too detailed. To fill this gap, PNNL is developing a testbed that integrates the miniWECC 240 bus power flow case within a new miniWECC market dataset. This talk will present the new 2032 miniWECC market dataset, which was augmented to reflect the capacity additions of the 2032 WECC Anchor Dataset, under the framework of the needs of future market models.

TB75

Regency - 706

Analytical Methods for Sustainable Manufacturing Planning

Invited Session

ENRE: Environment and Sustainability

Chair: Sidi Deng, University of Michigan, Ann Arbor, MI, 48108-2918, United States

1 - A Dynamic Material Flow Simulation Model for Risk-Informed Decision-Making in Decarbonizing Aluminum Manufacturing

Sidi Deng, University of Michigan, Ann Arbor, MI, United States, Daniel Cooper, John Sutherland

Aluminum is the world's second most consumed metal, and its production contributes substantially to global greenhouse gas (GHG) emissions. When formulating decarbonization strategies, it is imperative to ensure their coherence and alignment with existing industrial practices and standards. To support risk-informed decision policymaking in decarbonizing aluminum manufacturing, this study develops a dynamic system model that simulates global aluminum flows and computes their embedded GHG emissions. A baseline scenario is devised to reflect the current business and operation landscape, and three decarbonization strategies are proposed. Deterministic simulation is performed to generate dynamic material flows and performance metrics. Monte Carlo simulation is then implemented to evaluate the robustness of the system's performance under demand uncertainties. The results reveal that material efficiency and post-consumer scrap recycling have profound carbon performance implications. Informed by simulation outputs, macro decarbonization guidelines are formulated for various criteria. The object-oriented programming framework that underlies the dynamic MFA may be integrated with network analysis, agent-based simulation, and geospatial interfaces, which may lay the foundation for modeling more fine-grained material flows and supply chain structures.

2 - Digital Product Passports in the U.S.

Kelsea Schumacher, NIST, Gaithersburg, MD, United States, Noah Last, KC Morris

Businesses and policymakers are working to transition away from a linear—take, make, waste—economy into a circular economy (CE) model in which materials are kept in the economy for as long as possible through value-retention strategies like extended use, reuse, and recycling. However, this transition requires a system that shares product lifecycle data to support these value-retention strategies. A Digital Product Passport (DPP) has been proposed, and is being mandated for select products in Europe, as a mechanism to support such data sharing and drive circularity. The DPP concept relies on unique identifiers and decentralized ledger systems to provide extensive data throughout a product's lifecycle in a digital format that is both secure and accessible to some number of parties or stakeholders within the supply chain. Authorized stakeholders would then have access to select data, allowing for informed decision-making based on the product's history, content, or format. DPPs could support verifiable track and trace of materials as well as associated market surveillance. In this presentation, we will discuss the potential role and feasibility of using DPPs in the U.S., highlighting two use cases: batteries and textiles. We will discuss barriers facing the development and adoption of DPPs broadly, specific challenges and opportunities that may exist in the U.S., and preliminary work on the standards and frameworks required to ensure confidence in the quality and reliability of any information contained in them.

3 - Benchmarking Industrial Symbiosis Development using Network Analysis

Abheek Chatterjee, University of Maryland, College Park, MD, United States, Oren Minsk, Matthew Triebe, Buddhika Hapuwatte, Daniel Kietzer, Sushma Kittali-Weidner, KC Morris, Nehika Mathur

Industrial Symbiosis (IS) is the use of resource exchange partnerships between co-located firms to reduce raw material use, emissions, and waste generation, while promoting economic growth. Successful IS efforts contribute towards the Circular Economy transition. However, most IS projects fail in the early stages and do not reap the full sustainability benefits of IS. A significant reason for the failure of early-stage IS projects is the lack of knowledge regarding the strategic organization of IS networks. Quantifying IS network organization and operational characteristics and benchmarking IS development can address this challenge and guide decisions to advance newer IS networks by comparison to the organizational and operational characteristics of mature IS networks. This investigation tests multiple Ecological Network Analysis (ENA) and Social Network Analysis (SNA) metrics using three real-world IS case studies, as potential IS benchmarking tools. The results of the investigation show that ENA and SNA metrics can be used to compare IS networks and identify differences in the following characteristics: Partnership prevalence, producer-consumer relationships, linkage distribution, resource utilization, effect of indirect firm relationships, network cohesion, and network centralization. The values of the multiple network analysis metrics are found to be aligned with the development time of the IS networks and the qualitative markers of IS development stages as identified in the literature. This investigation demonstrates that network analysis of IS projects only requires anonymized and non-proprietary data, and identifies data-collection and modeling standards needed to develop an IS benchmarking framework using network analysis.

4 - The role of material scarcity in clean energy transition: diffusion control and circularity

Emre Nadar, Bilkent University, Ankara, Turkey, Atalay Atasu, Nilsu Uzunlar

The material scarcity may act as an important impediment to clean energy transition. We consider the problem of maximizing renewable energy generation over time to achieve the global climate goals under material scarcity. We model the market adoption of renewable technologies through a Bass diffusion curve and the decision maker's control over the diffusion curve through dynamic production volume choices. We explore novel trade-offs between renewable energy efficiency improvements, renewable technology diffusion growth and replacements, and circularity aimed at mitigating the material-scarcity issue. We identify conditions where delaying renewable technology adoption helps improve renewable energy generation in the long run. We also evaluate how the infrastructural limits to circularity affect the value of managing the diffusion process and to what extent circularity and diffusion control are complements. Our results show the importance of properly accounting for the diffusion process in clean energy transition.

5 - Using Reinforcement Learning to Manage Supply-Demand Balance in Remanufacturing

Abhijit Gosavi, Missouri University of Science and Technology, Rolla, MO, United States, Jiaqiao Hu

In this research, we propose a new reinforcement learning (RL) algorithm rooted in the **actor-critic** framework for approximately solving continuous state/action Markov decision processes. The algorithm uses a Q-function for the critic, which is in contrast to the usual value

function of dynamic programming used in many actor-critics. In this talk, we focus on a discretized version of this RL algorithm that uses exemplars or representative state-action pairs. This version is suitable for a tabular setting, and the algorithm is referred to as the *T*abular *E*xemplar *A*pproximation (*TEA*) version of the actor critic. In many remanufacturing firms, a key objective is to match the supply and demand imbalance commonly experienced with the so-called *cores* or used products. Suppliers of cores are in high demand, but their availability is limited. As a result, suppliers who deliver at more regular and reliable rates tend to be more expensive, while the more erratic suppliers are cheaper. Simultaneously, remanufacturing firms experience variability in their own customer base, and hence demand for their own products can fluctuate significantly. Therefore, unless the supplier is selected judiciously, firms can witness shortages in their finished-product inventory, leading to penalties from customers. The RL algorithm can streamline the supplier selection process, thereby improving efficiencies for the firm. Numerical results are presented to demonstrate the usefulness of the RL algorithm.

TB76

Regency - 707

Energy Systems Modeling and Optimization I: Policy-Enabling Models in the Energy Sector

Invited Session

ENRE: Other Energy

Chair: Afzal Siddiqui, Stockholm University, Kista, Sweden

1 - Treading Water on a Cloudy Day: Utility Rate-Making in the Presence of Prosumers

Yihsu Chen, University of California, Santa Cruz, Santa Cruz, CA, United States

PUCs (Public Utility Commissions) in some states in the U.S. face daunting challenges in making equitable and efficient retail rates while adequately recovering utilities' fixed costs given a growing amount of DER (distributed energy resources) dominates the grid. We examine the trade-offs among energy equity, economic efficiency, and utility cost recovery in retail ratemaking by formulating the problem as a two-stage stochastic problem. A conditional-value-at-risk approach is implemented to account for the PUC's attitude toward recovering utilities' fixed costs. We demonstrate the (im)possibility of devising a retail rate that is equitable while also minimizing the risk of unrecovered fixed costs at the intangible expense of efficiency. Our work, thus, highlights the need to strike a delicate balance among those factors in today's ratemaking.

2 - A Risk-Averse Multi-Stage Stochastic Programming Approach to Generation and Transmission Expansion Co-Planning

Sebastian Maier, University College London (UCL), London, United Kingdom, Yifeng Yu, Tohid Erfani

In the ever changing landscape of the electricity market, proper alignment for generation and transmission capacity expansion planning is crucial to meet future demand, mitigate financial risk and maintain grid stability. Traditional models often misestimate the risks involved, leading to potentially suboptimal investment decisions. Integrating risk aversion measures, such as CVaR (Conditional-Value-at Risk), into the planning process is a complex challenge. It requires ensuring time consistency in the planning process while trading-off an efficient capacity expansion decision. This paper proposes a time consistent multistage generation and transmission expansion co-planning model. The model employs a nested formulation of CVaR as a risk assessment metric and the economic impact of the planning decisions as the financial measure. The model is applied to the New England IEEE 39 bus test system to examine the impact of risk aversion on the strategies for time consistent optimal capacity expansion planning in a perfectly competitive electricity market. The results present the strategy transition in terms of different levels of risk aversion on optimal expansion decisions. These shifts in investment behaviour are consistent with practical considerations and real-world scenarios, emphasising the importance of appropriately considering risk aversion in electricity capacity expansion planning models. We compare time consistent and inconsistent models in the same settings to understand how risk measures impact investment decisions. The time consistent model results in lower investment levels, suggesting more conservative choices. This research has implications for policy-making, highlighting the importance of a nuanced understanding of risk in planning.

3 - Managing Basin-Scale Carbon Sequestration: a Tragedy of the Commons Approach

Joseph Duggan, University of Dayton, Dayton, OH, United States, Jonathan Ogland-Hand, Michael Hannon, Steven Anderson, Richard Middleton

Reaching net-zero will require gigatonne-levels of CO_2 sequestration. For example, recent models of the energy system in the United States suggest that up to two gigatonnes of CO_2 storage may be required in 2050. This amount of CO_2 injection is multiple orders of magnitude greater than current injection rates can achieve and will require a huge expansion of deployment of CCS that could lead to competition between firms that sequester CO_2. For example, given that formations within sedimentary basins generally allow lateral fluid flow, the induced pressure increase from one firm's CO_2 injection may limit another firm's ability to inject CO_2. In this way, basin-scale carbon sequestration has the two main hallmarks of a classic Tragedy of the Commons, where the storage resource is 1) non-excludable, in that multiple CO_2 storage operators may be permitted to inject into the same formation without overlapping CO_2 plumes but still cause pressure interference with each other, and 2) rival in consumption, meaning each firm's use of the geologic subsurface for sequestration purposes reduces other firms' ability to profitably sequesterCO_2 because of the increased pressure. Here, we explore the incentive implications of basin-scale carbon sequestration by modelling geologic CO_2 storage as a variant of the Tragedy of the Commons, a well-studied problem in the economic literature. CO_2 capture and storage (CCS) is an essential tool for meeting global climate goals, but the incentive effects, market design implications, and broader economic management strategies of CCS are underexplored.

4 - Day-ahead and reserve prices in a renewable-based power system: adapting electricity-market design for energy storage

Laureen Deman, Univ. Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL, Grenoble, France, Afzal Siddiqui, Cédric Clastres, Quentin Boucher

Decarbonising the power mix will require investments in storage and flexibility options to replace the current carbon-intensive supply of reserves. This paper questions whether reserve-capacity markets can serve as a capacity mechanism for flexible technologies. A fundamental model of the day-ahead and reserve markets is used to investigate the evolution of reserve prices with large shares of renewable energy and storage. The model represents the current market design in Continental Europe with a centralised supply and platforms for the exchange of reserves. By becoming the main suppliers of reserve capacity, batteries have a noticeable impact on reserve prices. Their flexibility implies zero opportunity cost most of the time, meaning that the flexibility is not rewarded by the market. These results suggest that reserve-capacity markets cannot provide additional remuneration for flexible technologies and, thus, do not solve the missing-money problem in the context of the energy transition.

5 - Empowering Energy Communities: Balancing Technological, Economic, and Social Objectives through Optimization

Lia Gruber, Graz Unversity of Technology, Austria, Graz, Austria, Sonja Wogrin

Facilitating the transition to sustainable energy systems necessitates the active involvement of ordinary citizens, which can be achieved through energy communities. These communities enable collective energy production, consumption, storage, and trading, serving as multi-dimensional constructs involving various stakeholders. While often approached solely from a technological standpoint, energy communities inherently carry social and economic dimensions, impacting issues like energy poverty. To address these complexities, we propose a comprehensive optimization model that considers all dimensions through a multi-objective approach. Our model aims to optimize three key objectives. Firstly, it seeks to minimize power peaks within the energy community, minimizing strain on the grid. Secondly, it aims to minimize the overall cost of the energy community, considering factors such as operational cost, grid tariffs and energy purchasing. Thirdly, it prioritizes maximizing fairness for energy-poor members within the community, ensuring equitable access to affordable energy resources. It integrates diverse production and flexibility options, including biomass, photovoltaic, battery storage, and electric vehicles, while also incorporating grid constraints and tariffs as well as energy allocation between members. Grid tariffs and energy allocation methods implemented in the model are accordance with Austrian legislation. As implementation, subsidization, and regulation of energy communities have already been well established in the country. By optimizing trade-offs between technological, economic, and social objectives, our model seeks to guide the design and operation of sustainable and inclusive energy communities, advancing the transition towards a more decentralized and equitable energy landscape.

TB77

Regency - 708

AI and Optimization for Decision-Making

Invited Session

Computing Society

Chair: Leonard Boussioux, UW Foster, Harvard, Seattle, WA, 98122, United States

1 - Climate Smart Staffing for Harvesting Operations

Sanchita Das, University of Washington, Seattle, WA, United States, Masha Shunko, Leonard Boussioux

Problem Definition: We present our collaboration with the Washington Fruit Tree Research Center to design a capacity planner that enables dynamic resource allocation for fruit tree growers in response to weather and climate based yield uncertainties to maximize farmer's profit. Apple production is a 4 billion USD industry, and labor contributes 25-40% of input costs. Given the fluctuations in weather and climate due to global warming, the time of apple ripening (harvest window) is becoming erratic from year to year. The pre-season planned resource (labor, refrigeration) capacity is either under or over-utilized. We contribute a robust optimization technique that calculates and dynamically allocates labor to farms by predicting shifts in harvest dates due to weather changes. Methodology/Results: We combine a weather forecasting model with a climate adaptation model for apple fruit trees to predict fruit maturity (harvest) dates. We model a data-driven dynamic staffing optimization and robustify it using a distributionally robust optimization where ambiguity sets are determined based on the predictions from historical data. Managerial implications: Due to global warming, it is imperative that growers have a set of decision support tools to hedge against climate change. Our proposed capacity planning tool provides early alerts to farmers for dynamic staffing in response to weather based uncertainties to improve their preparedness and profitability.

2 - Data-Driven Stochastic Programming for Electric Vehicle Charging Station Location with Decision-Dependent Demand Learning

Huangrong Sun, The Ohio State University, Columbus, OH, United States, Xian Yu, Guzin Bayraksan

We consider a two-stage electric vehicle charging station location problem with uncertain customer demand, where the demand depends on the first-stage location decisions. We employ both parametric (e.g., exponential and linear regression) and nonparametric (e.g., kNN) models to describe the dependence between customer demand and location decisions and develop a two-step learning approach to fit the regression model. Using empirical residuals from the learning step, we construct a residuals-based sample average approximation problem to approximate the true two-stage stochastic program. Numerical studies are conducted on synthetic data and real-world data in New York State to evaluate result sensitivity and compare different approaches.

3 - Search Between the Lines: Field Experiment, Structural Estimation, and LLM-Based Optimization

Jin Liu, University of Science and Technology of China, Hefei, China, People's Republic of, Xingchen Xu, Yongjun Li, Yong Tan

In online marketplaces, consumers are typically presented with lists of products, allowing them to browse and search items of interest for further information before finalizing their purchases. Existing literature has explored consumer interactions with vertically arranged product lists; however, empirical evidence is lacking for horizontal lists, another prevalent format. Moreover, in specific sectors like online reading, text elements can significantly influence both the search costs and the expected utility derived from a product, complicating the analysis of these dual effects. To address the gap, we develop a sequential search model that integrates considerations of horizontal ranking, text-associated search costs, and text-influenced utility expectations. Utilizing data from a field experiment conducted on a prominent online reading platform, where books are displayed in a randomized horizontal order, we estimate our model to examine how horizontal rankings and text features impact consumer search behaviors and decision-making processes. Additionally, we propose a counterfactual policy, utilizing large language models, to enhance decision-making efficiency in such settings.

TB78

Regency - 709

Computational Integer Programming

Invited Session

Computing Society

Chair: Chen Chen, The Ohio State University, Columbus, OH, United States

1 - Two-Column Probing for Mixed-Integer Programming

Chen Chen, The Ohio State University, Columbus, OH, United States

Probing involves tentatively testing values of variables in order to determine various implications. In mixed-integer programming probing is typically conducted on one binary variable at a time in the preprocessing stage, yielding implications that are useful for branch-and-cut algorithms such as conflicts, strengthened variable bounds, aggregations, and cutting planes. In this talk we consider the extension to two-column probing, where pairs of binary variables are tested in tandem (over their four possible configurations). Although such a procedure may yield more and stronger implications compared to standard probing, the computational load is substantially higher. We describe our attempt to navigate this tradeoff.

2 - Exploiting Symmetry in Risk-Averse Antibiotics Time Machine Problem

Burak Kocuk, Sabanci University, Istanbul, Turkey, Deniz Tuncer

Antibiotics time machine problem is a challenging combinatorial optimization problem that can be used to reverse the antibiotic resistance of the bacteria. The risk-averse version of this problem can be modeled as a mixed-integer linear program with scenarios and can be solved by adapting a scenario decomposition approach, which is based on no-good cuts. We develop stronger versions of the standard cuts that get rid of several solutions at the same time with few inequalities. We also study the underlying polyhedral set that involves the so-called symmetry breaking constraints and show how the aforementioned cuts can be further improved. Finally, we present the results of our computational experiments involving a real dataset to illustrate the efficacy of our approach.

3 - Augmenting Seq2Seq Models with Constraint Reasoning for Ambiguous Combinatorial Optimization

Macarena Navarro, Carnegie Mellon University, Pittsburgh, PA, United States, Willem-Jan Van Hoeve

We consider combinatorial optimization problems with ambiguous objectives or constraints, for which the functional form is not known. Learning the ambiguity in such problems is a generalization of inverse optimization. We propose to represent the ambiguous components using a neural network. While these could be linearized and added to an optimization model, that approach is typically not scalable. Instead, we propose using sequence-to-sequence models that learn from past decisions and, given an input specification, can directly generate similar solutions to solve the problem. To ensure feasibility of the generated solutions we embed a constraint reasoning layer that represents the known hard constraints of the model. Using a knapsack problem with unknown objective as a case study, we demonstrate that our approach can generate near-optimal solutions and can even outperform an omniscient greedy heuristic.

4 - Can an infeasible MIP solve itself?

Yuriy Zinchenko, University of Calgary, Calgary, AB, Canada

The analysis of why a specific MIP instance is infeasible formally can be reduced to computing an Irreducible Infeasible Subset (IIS) of the constraints. Unlike the case of LP, for MIP there is no useful duality that can be employed to facilitate such computations. The process of determining an IIS for MIP is typically handled with brute force, e.g., by use of deletion filters and alike, thus rendering IIS determination for a MIP into a much harder computational task. We will discuss one approach to optimizing this process and what components of this approach have made it into the newest version of Gurobi.

Tuesday, October 22, 12:45 PM - 2:00 PM

TC01

Summit - 320

Empirical Research in Service Operations

Invited Session

Service Science

Chair: Jane Iversen, The Ohio State University, Columbus, OH, United States

Co-Chair: Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - When does Continuity matter? Evidence from a health services setting.

Vishal Ahuja, Southern Methodist University, Dallas, TX, United States

The topic of continuity of care (CoC) has gained attention within the operations literature, especially recently. Studies have generally found a positive association between (greater) primary care CoC and (improved) outcomes such as 30-day readmissions. However, limited attention has been paid to how CoC interacts with other operational levers – that is, whether and how the benefits of continuity alter in the presence of other operational levers that managers may have at their disposal. Thus, we first study whether the benefits of CoC extend to mental healthcare. Importantly, we investigate continuity's relationship with four operational levers that have typically been studied independently– collocation of physical and mental health services, visit regularity, team familiarity and patient complexity. As America's largest integrated health system, the Veterans Health Administration offers an ideal setting to investigate these questions. We analyze more than 73,000 patients – over an 11-year period – who suffer from chronic conditions and show evidence of mental illness. We find that CoC is associated with improvement in hospitalizations, length of stay (LOS), 30-day readmits, and suicidal behavior. For example, 1 StDev increase in CoC is related to 8.2% average reduction in LOS, equivalent to $7 million annually. We also find that continuity has a substitutive relationship with collocation, familiarity, and regularity, but a complementary relationship with complexity (that is, the benefits of CoC are further enhanced for severe patients). Our work offers practical insights into where to target CoC and, broadly, how to design an operationally efficient system.

2 - Revving Up Retention: Analyzing the Influence of the Introduction of Food Delivery Services on Driver Behavior

Michael Dwyer, Iowa State University, Ames, IA, United States, Sukrit Pal, Thu Trang Hoang, Chris Tang

To sustain growth amidst increasing competition, ride-hailing platforms often introduce adjacent services, like food delivery, to their mobile apps for customers and drivers. However, the impact on drivers' choices between ride-hailing and new adjacent services remains understudied, motivating this project. By analyzing data from a ride-hailing platform in Vietnam, where a new food delivery service was introduced in two cities at different times, we used difference-in-difference with robustness checks to examine drivers' choices. Our analysis found that hybrid drivers, those offering both ride-hailing and food delivery services, saw an increase in revenue and worked more hours. We also found that these effects varied between part-time and full-time hybrid drivers. Hybrid drivers opted for shorter rides, clustering their operations. Meanwhile, dedicated drivers providing only ride-hailing services chose longer rides. The introduction of the adjacent food delivery service reduced labor supply disparities between part-time and full-time drivers. However, our study indicates "cherry-picking" behaviors among hybrid drivers with shorter rides, while dedicated drivers demonstrated a preference for longer rides. The adjacent service introduction created a potential segmentation in the driver pool and created new risks to service operations for the platform.

3 - Multi-Channel Healthcare Operations: The Impact of Video Visits on the Usage of In-Person Care

Tan (Suparerk) Lekwijit, W. P. Carey School of Business, Arizona State University, Tempe, AZ, United States, Hummy Song, Christian Terwiesch, Krisda Chaiyachati

Healthcare organizations have increasingly adopted video visits as an alternative care channel that offers greater convenience and holds promise for improving patients' access to care. Despite the growing use, the impact of a new digital channel on care demand within existing physical channels has not been thoroughly investigated, especially in the primary care context of a traditional healthcare organization. We study a large healthcare system that made video visits available to a subset of its patients for primary care needs, prior to the COVID-19 pandemic. Using a difference-in-differences approach and visit data spanning over 9.5 million patient-months, we find that the introduction of video visits increases the demand for in-person primary care provider (PCP) visits by 20% and the demand for emergency department (ED) visits by 30%. When considering only patient-initiated visits, we find that the introduction of video visits increases the overall number of care initiations but does not affect in-person care initiations, suggesting that patients initiate care through video visits for care needs that may have otherwise gone unattended. Furthermore, we find that the increase in in-person demand arises primarily from patients who have poorer access to in-person care, as measured by distance and provider busyness. Lastly, compared to in-person visits, we find that video visits are less likely to provide care that can successfully conclude a care episode, especially when the medical condition is acute. Our findings have important capacity and revenue implications for healthcare organizations that are considering offering video visits as an additional care channel.

TC02

Summit - 321

Computational/Statistical Methods for Uncertainty Quantification: Part II

Invited Session

Quality, Statistics and Reliability

Chair: Moses Chan, Northwestern University, Evanston, IL, United States

Co-Chair: Ozge Surer, Miami University, Oxford, OH, United States

1 - Modeling High-Dimensional Outputs with Heterogeneous Noise Structure

Moses Chan, Northwestern University, Evanston, IL, United States

Computer models, or simulators, are indispensable in studying physical systems. Modern computer models generally are time-consuming to run and potentially produce noisy high-dimensional outputs. The calibration of such a computer model is facilitated by the use of fast surrogate models. This talk presents the latent component Gaussian process that leverages a basis representation of the output and models each latent component as a Gaussian process. The resulting formulation admits a general error covariance matrix of the noise, extending previous works which assume homogeneous error variance across the outputs. With an appropriate linear transformation of the output, this method provides exact inference in its prediction without incurring significant additional computational costs. We demonstrate the strength of this method through a set of numerical experiments and a case study involving the emulation of a viscous anisotropic hydrodynamics model used in heavy-ion collision simulations.

2 - Co-Active Subspace Methods for the Joint Analysis of Adjacent Computer Models

Kellin Rumsey, Los Alamos National Laboratory, Chicago, IL, United States

Active subspace (AS) methods are a valuable tool for understanding the relationship between the inputs and outputs of a Physics simulation. In this paper, an elegant generalization of the traditional ASM is developed to assess the co-activity of two computer models. This generalization, which we refer to as a Co-Active Subspace (Co-AS) Method, allows for the joint analysis of two or more computer models allowing for thorough exploration of the alignment (or non-alignment) of the respective gradient spaces. We define co-active directions, co-sensitivity indices, and a scalar "concordance" metric (and complementary "discordance" pseudo-metric) and we demonstrate that these are powerful tools for understanding the behavior of a class of computer models, especially when used to supplement traditional AS analysis. Details for efficient estimation of the Co-AS and an accompanying R package (concordance) are provided. Practical application is demonstrated through analyzing a set of simulated rate stick experiments for PBX 9501, a high explosive, offering insights into complex model dynamics.

3 - Rational Kriging

Roshan Joseph, Georgia Tech, Atlanta, GA, United States

I will talk about a new kriging method that has a rational form. It is shown that the generalized least squares estimator of the mean from rational kriging is much more well behaved than that of ordinary kriging. Parameter estimation and uncertainty quantification for rational kriging are proposed using a Gaussian process framework. I will also discuss a generalized version of rational kriging, which includes ordinary and rational kriging as special cases. Extensive simulations carried out over a wide class of functions show that the generalized rational kriging performs on par or better than both ordinary and rational kriging in terms of prediction and uncertainty quantification. The potential applications of the new kriging methods in the emulation, calibration, and optimization of computationally expensive models will be illustrated with real and simulated examples.

TC03

Summit - 322

Data Science and AI for Informing Decisions in Mission Critical Systems

Invited Session

Quality, Statistics and Reliability

Chair: Mingyang Li, University of South Florida, Tampa, FL, United States

Co-Chair: Qingpeng Zhang, City University of Hong Kong, Kowloon, Hong Kong

1 - A Reinforcement Learning Approach to Network-level Maintenance Planning of Co-located Road and Water Infrastructures with Interdependencies

Bingjie Wang, UNIVERSITY OF SOUTH FLORIDA, Tampa, FL, United States, Hung Nguyen, Qing Lu, Qiong Zhang, Mingyang Li

The complex physical and operational interdependencies between co-located pipes and roads pose significant challenges in determining optimal maintenance actions for the deteriorating water infrastructure (WI) and transportation infrastructure (TI). For example, traffic load could exacerbate the co-located pipe failure rates, and the pipe's failure could in turn cause the blockage of its co-located roads, potentially leading to widespread traffic disturbances, incurring high corrective maintenance costs, user costs, and traffic control costs. Many existing maintenance works focus on studying a single system, such as WI or TI, and there is a lack of research to investigate the joint maintenance of co-located WI and TI with complex interdependencies. In this study, we propose a network-level maintenance decision-making framework for jointly optimizing the maintenance actions of TI and WI, consisting of a large number of co-located roads and pipes with both physical and operational interdependencies. Given the large action space inherent in network-level decision-making, a reinforcement learning approach is considered to determine the optimal maintenance actions across the networks, aiming to enhance cost efficiency and infrastructure robustness. To illustrate the proposed work and demonstrate its performance, a case study based on the co-located TI and WI in Ybor City, Tampa, is provided to showcase the cost-efficiency of the proposed maintenance strategies as compared to some existing benchmarks that neglect the complex interdependencies between TI and WI.

2 - A Deep Learning Enhanced Probabilistic Approach to Predicting Discharge Outcomes of Post-acute Care Residents in Skilled Nursing Facilities

YULUN XU, University of South Florida, Tampa, FL, United States, Nazmus Sakib, Hongdao Meng, Nan Kong, Kelly Smith, Mingyang Li

Predicting the discharge outcomes and their timings of the post-acute care (PAC) residents from skilled nursing facilities (SNFs) is essential for improving quality care, maintaining residents' functional status, and enhancing their chances of successfully returning to the community, private homes or lower-level care settings. However, it could be challenging due to random SNF dwelling duration with multiple discharged dispositions (e.g., community, re/hospitalization), as well as the influence of varied individual characteristics, diseases, and functional performance.

Existing work mainly focused on predicting a single discharge outcome and its timing in the hospital environments, and many of existing prediction models can only provide deterministic prediction outputs. We propose a deep learning enhanced probabilistic prediction model which is capable of accurately predicting resident-level time-to-discharge outcomes with multiple discharge dispositions. The proposed model takes advantage of deep learning’s capability of establishing nonlinear functional mapping and further allows the probabilistic assessment and prediction of time-to-discharge outcomes with meaningful model interpretations. A real case study is provided to illustrate the proposed work and demonstrates its superior performance (e.g., prediction performance, model interpretation) over existing benchmarks.

3 - Optimal Bidding in Online Display Advertising: A Deep Reinforcement Learning Model for Mobile Gaming Market

Mengzhuo Guo, Sichuan University, Chengdu, China, People's Republic of, Qirui Yang, Congde Yuan, Qingpeng Zhang, Frank Chen

When a user opens a website or app, there are often several advertisements interspersed. Whenever such an advertisement appears, loads, and is seen on a user's screen, it is referred to as an impression. For advertisers, an impression signifies an opportunity for their ads to be viewed. Real-time bidding (RTB) has emerged as a prominent paradigm, allowing advertisers to procure available impressions through instantaneous automatic auctions. In RTB, the goal for advertisers is to maximize the total revenue generated from the impressions they win while considering constraints such as budget. Due to the difficulty in estimating the value of impressions and the bids of competitors, determining the optimal bid for each impression becomes a highly challenging problem for advertisers. To address the sparse reward issues causing an unstable learning process, we first propose a novel definition of impression value under the background of the online mobile gaming market, which takes into account players' in-game purchase tendencies, conversion revenue, and publisher characteristics. In addition, to enhance the interpretability of the strategy, we divide the bidding process into two parts: the first part employs reinforcement learning techniques to obtain a two-dimensional vector based on an optimization problem. This vector contains information on both the risk of overpaying and the likelihood of securing a bid. We conduct comprehensive offline experiments and set up an online A/B testing experiment. Real-world experiments with more than a hundred million users during a week validated that ACRL improved the return on investment by 83.89%.

TC04

Summit - 323

Optimization for Smart Grids

Contributed Session

Chair: Shahab Sadri, University of Louisville, Louisville, KY, United States

1 - A Real Options Valuation to Commercialize Wave Energy Farms

James DiLellio, Pepperdine Graziadio Business School, Malibu, CA, United States, Effie Ashu

Waves are one of the most abundant sources of renewable energy, yet the technology needed to profitably harvest it remains nascent in the commercial markets. Wave energy is obtained by converting kinetic energy from ocean waves into electricity. One of the big challenges in commercializing wave energy lies in the uncertain investment needed to increase the Technology Readiness Level (TRL) to 9, the level at which technology is considered ready for widespread deployment in the market while properly addressing the economic and technical challenges of energy production at scale. Consequently, to analyze these investments we use the concepts of real options, treating the investment as an out-of-money compound and deferred option starting at TRL 6. A real option valuation is conducted on these risky investments amid uncertain wholesale electricity prices. We show that the real option results may turn a negative Net Present Value (NPV) project without managerial flexibility into a positive one due to the expected option value. Additionally, we show that with sufficient carbon credits as a percent of capital expenditures, the project becomes even more favorable for investment and reduces the risk of negative NPV further. The results of this research support government decisions on the need for additional investment and/or incentives for renewable energy projects. We also suggest the possibility of public-private partnerships for any nascent technology like wave energy that still requires additional investment in research and development to reach TRL 9.

2 - Impact of Climate Change on Battery Energy Storage: A Case Study of MISO

Zhenggao Wu, University of Waterloo, Waterloo, ON, Canada, Stanko Dimitrov, Michael Pavlin

Climate change will have multiple impacts on energy markets by changing demand patterns and changing supply patterns --- by indirectly increasing renewable penetration and changing the dynamics of the supply from renewable sources (e.g., when and where it is windy/cloudy). Battery energy storage systems (BESSs) are one way to effectively manage intermittency from renewable energy. However, their profitability and incentive to participate in markets are susceptible to both the magnitude and frequency of price variation. This paper investigates the impact of climate change on a BESS operating in a North American deregulated electricity market. We propose a robust optimization model to determine the operating policy for a BESS over an 80-year period, considering various climate projections. We reformulate the robust optimization model to an equivalent linear program that allows us to numerically explore 108 climate scenarios. Our empirical study focuses on the Midcontinent Independent System Operator (MISO) market in the U.S., aiming to determine the optimal operation of the BESS using our model. The findings reveal that precipitation is the primary driver of variability in electricity prices, which varies across climate scenarios and geographical locations. Furthermore, climate change could increase arbitrage opportunities and operational incomes through arbitrage in Northern MISO, likely due to the region's heightened sensitivity to wind variations.

3 - Fair Control Policy for Shared Battery Energy Storage

Chaiyoung Lee, Yonsei University, Seoul, Korea, Republic of, Soongeol Kwon

With the increasing interest in improving renewable energy penetration and sustainability, Battery Energy Storage (BES) has been widely adopted. Shared BES, in particular, offers advantages in terms of operational and cost efficiency compared to individual BES, and thus, the installation of shared BES has been accelerated. However, to promote and sustain the shared BES, it is essential to ensure fair access and guarantee the financial benefits of sharing BES across users. This study aims to design a control policy that allows users to share BES with fairness and benefits by incorporating the uncertainties and variability existing in solar energy generation, electricity demand, and electricity prices. Specifically, a two-stage stochastic mixed linear programming is proposed to derive equitable charging and discharging thresholds that can be used as a rule-based control policy. The proposed control policy is validated and evaluated through comprehensive numerical experiments where the findings emphasize its capability and practicability to enhance the operations of shared BES.

4 - Managing residential electricity consumption with buyback price in a smart micro grid

Shahab Sadri, University of Louisville, Louisville, KY, United States, Lihui Bai

This research aims to study the residential electricity load in a smart grid consisting of a microgrid with solar panels and battery storage systems under the net metering program. The system allows residents to sell their excess electricity back to the grid. We study the smart grid from the perspectives of households as well as the utility company. In the Social Welfare model (SW) model, the utility company aims to maximize the social welfare of the entire network, which includes the households' convenience minus the electricity generation cost. On the other hand, in the User Equilibrium model (UE) model, each household maximizes his or her utilities in a Nash Equilibrium game. In the real world, the users' consumption is affected by the pricing scheme that the utility company charges. Therefore, we model the hierarchy as a Price-Responsive User Equilibrium (PRUE) model. In the PRUE model, each household maximizes its utilities in response to the pricing scheme imposed by the utility company. We show that alternative pricing schemes can achieve the maximum possible social welfare and, therefore, can be further optimized using secondary objectives. The latter results in different users' consumption behavior. Numerical results with various pricing schemes and sensitivity analysis are to be reported.

TC05

Summit - 324

ESG Strategies and Corporate Performance

Contributed Session

Chair: Yang Li, Montclair State University, Montclair, NJ, United States

1 - Heterogeneity Aspiration Performance Gaps and Corporate Environmental Responsibility: A Comparison between Family Firms and Non-family Firms

Xuan Wang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Wei Sun

Fulfilling environmental responsibility is an important practice for corporate transformation and upgrading, as well as a crucial strategic decision for businesses. However, existing literature lacks analysis from the perspective of performance feedback in examining the key factors that drive companies to fulfill environmental responsibilities. Based on performance feedback theory, this study explores the impact of heterogeneous aspiration performance gaps on corporate environmental responsibility and compares the similarities and differences in this relationship between family firms and non-family firms. Using data from Chinese A-share listed companies from 2009 to 2022, the study finds that historical aspiration performance gaps and social aspiration performance gaps have different impacts on corporate environmental responsibility. When distinguishing between types of companies, family firms demonstrate different moderating effects on this relationship. This study expands the understanding of how aspiration performance gaps influence corporate environmental responsibility and provides valuable insights for companies in formulating their environmental responsibility strategies, especially regarding the differentiated approaches that should be taken based on the type of enterprise.

2 - Do Well by Doing Good? The Trend and the Operational, Market Valuation, and Employee Welfare Implications of Rank-and-File Employee Ownership Disclosure

Runqing Yang, Boston University, Boston, MA, United States, Mohamed Balboula, Nuo Yuan, Jack Moriarty, Dokyun Lee

Employee ownership (EO) is increasingly recognized for its potential to enhance firm performance, employee well-being, and innovation. Despite these advantages, a critical gap remains in understanding how public companies disclose rank-and-file EO and how such disclosures impact key stakeholders, particularly investors, employees, and customers. This gap is even more pronounced in light of recent regulations like the SEC's 2020 human capital disclosure rule. This study directly addresses this gap by analyzing EO disclosures in 10-K filings of U.S. public firms implementing NCEO-validated ESOPs from 2010 to 2023. We employ a multi-step approach. First, we utilize keyword-based text extraction to isolate relevant sections within extensive corporate reports focusing on EO terms. Next, we leverage advanced large language models (LLMs) to construct various EO disclosure indexes, quantifying the quality and depth of corporate disclosure. Finally, our econometric analysis, accounting for potential LLM-generated measurement errors, explores how these disclosure indices correlate with market valuation (stock returns and risk), operational performance (profitability and productivity), and employee welfare (measured through Glassdoor ratings and turnover rates). This research not only extends the knowledge base on corporate disclosure and human resource management but also provides actionable insights. It can help corporate managers and policymakers leverage EO disclosure practices not just for compliance but as a strategic asset to boost firm performance and stakeholder engagement, particularly in attracting and retaining talent, and influencing investors and customer behavior.

3 - Analytics on Puzzling Relationship Between ESG and Corporate Financial Performance

Akira Maeda, The University of Tokyo, Tokyo, Japan, Hiroshi Ishijima

The impact of ESG (Environmental, Social and Corporate Governance) and CSR (Corporate Social Responsibility) on corporate performance has long been discussed. People are mainly interested in whether or not ESG helps improve corporate performance. There are many academic studies on this topic, but the results are not uniform: some studies show a positive relationship between the implementation of ESG activities and corporate financial and business performance indicators e.g., profit, revenue, return on assets (ROA), and return on equity (ROE). On the other hand, some studies argue that there is a negative relationship between ESG and corporate performance indicators. Whichever conclusion is reached, most of the research to date has been a discussion of whether the relationship is positive or negative, not the idea that it is essentially a mixture of both effects. The conclusion is a choice between the two.

This paper is an attempt to provide a new perspective for understanding these mixed results thus far. We propose the hypothesis of a U-shaped relationship as a hypothesis that can support both sides of the argument. The authors have reported on this hypothesis up to that point in several seminars and presentations including Maeda and Ishijima (2023a) as well as a short article of Maeda and Ishijima (2023b). In this paper, we report our latest findings, especially on the theoretical analysis: After testing out hypothesis using data from the Toyo Keizai CSR Ranking, which is a comprehensive survey report on Japanese companies, we present an innovative theory.

4 - Managing Sustainable Environment, Social, and Governance (ESG) Activities in Health Care Organizations Using Goal Programming Model

Chang Won Lee, Hanyang University Business School, Seoul, Korea, Republic of

This paper is to develop a goal programming (GP) model that aids in managing sustainable environment, social, and governance(ESG) activities in health care organizations. The model is developed on the ESG related data from health care organizations. The suggested model will utilize a GP model to reflect the multiple, conflicting ESG activities. The ESG activity goals are decomposed and prioritized and the model results and findings are discussed. This suggested GP model will provide strategic insights in health care ESG planning and similar planning settings.

5 - Attention Management by Salience in ESG Report

Yang Li, Montclair State University, Montclair, NJ, United States

ESG reports have become an essential part of corporate communication and serve as a tool for firm reputation and impression management. Based on the salience theory, this study employs a deep learning-based visual attention algorithm to analyze if and how firms with different levels of commitment to sustainability manage the readers' attention by manipulating bottom-up attention transmitted through environmental, social, and governance (ESG) reports. Our results show that reports from less committed firms are more likely to show distracting patterns with an overwhelming amount of salient points and regions populated across all pages. Stakeholders encountering reports with numerous distracting elements should exercise caution and suspect that the company may not accurately represent its commitment to ESG.

TC06

Summit - 325

Advances in Reinforcement Learning

Contributed Session

Chair: Mohammad Rostami, Alarm.com, Arlington, VA, United States

1 - Link Prediction Modelling in Human Trafficking Networks - A Survey

Hasini Balasuriya, University of Louisville, Louisville, KY, United States, Monica Gentili

Link prediction models play a key role in analyzing complex networks, forecasting missing or future connections, and revealing the dynamics of illicit trafficking networks. While numerous surveys have examined general link prediction modeling, research specifically exploring these models' application to illicit networks remains sparse. This study contributes a thorough literature review that addresses this deficiency by evaluating how effectively link prediction models can identify missing links in illicit networks. Our aim is to deliver a precise, well-organized overview that advances understanding of the potential of these models in disrupting illicit activities.

2 - State augmented reinforcement learning for constrained Markov games

Soham Das, Texas A&M University, College Station, TX, United States, Ceyhun Eksin

We present a new approach for solving constrained Markov games utilizing a state-augmented description of constrained reinforcement learning. We consider a Markov game with identically interested agents with local cost constraints that depend on the action and state coupled over time. These games, for example, serve as a formal framework to study autonomous teams with local energy and time constraints. We augment the state space of the Markov game with Lagrange multipliers and reinterpret primal dual methods as the portion of the dynamics that drives the multipliers forward. Our algorithmic framework assumes that agents can solve the unconstrained Lagrangian Markov game to an approximate Nash Equilibrium. Then our primal dual solver guarantees feasibility for the constrained Markov game asymptotically, which still converging to an approximate Nash equilibrium for the game.

3 - Integrating Reinforcement Learning and Stochastic Network Design

Fatemeh Sarayloo, University of Illinois Chicago, Chicago, IL, United States, Teodor Gabriel Crainic, Walter Rei

This research aims to explore the integration of reinforcement learning (RL) algorithms and optimization methods to address stochastic network design problems. The goal is to enhance the synergy between reinforcement learning and optimization methods to effectively explore extensive solution spaces. The anticipated outcomes promise to significantly influence the field by offering more robust and efficient solution methods for complex, large-scale optimization problems in uncertain environments.

4 - Learning Efficiently in Data-Scarce Regimes

Mohammad Rostami, Alarm.com, Arlington, VA, United States, Mohammad Rostami

The unprecedented processing demand, posed by the explosion of big data, challenges researchers to design efficient and adaptive machine learning algorithms that do not require persistent retraining and avoid learning redundant information. Inspired from learning techniques of intelligent biological agents, identifying transferable knowledge across learning problems has been a significant research focus to improve machine learning algorithms. In this talk, we explain how the challenges of knowledge transfer can be addressed through embedding spaces that capture and store hierarchical knowledge.

We first focus on the problem of cross-domain knowledge transfer. We explore the problems of zero-shot learning and unsupervised domain adaptation and explain how similar ideas can be used to address both learning settings. We then investigate the problem of cross-task knowledge transfer in sequential learning settings. We address the problem of lifelong machine learning using generative experience replay, where the goal is to learn sequential tasks without forgetting. We develop a method to overcome the problem of catastrophic forgetting within a continual learning setting of deep neural networks by enforcing the tasks to share the same distribution in the embedding space.

Finally, we focus on current research directions to expand the past progress and plans for the future research directions. Through this talk, we demonstrate that despite major differences, problems within the above learning scenarios can be tackled using a unifying strategy that allows transferring knowledge effectively.

5 - Learning Sentence Similarity Using Siamese LSTM Network and Traditional Approaches

Sajede Farahani, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Community Question Answering platforms such as Quora, Stack Overflow, and others offer users a venue to engage with experts and acquire precise and reliable answers to their inquiries. To mitigate the delay between posting a question and receiving a response, one strategy involves retrieving analogous past questions from the archives of these platforms. Identifying semantic similarity presents a complex challenge due to the vast array of potential expressions available in natural language, alongside its inherent ambiguity. This study adopts two approaches. Firstly, we utilize feature engineering alongside various machine learning algorithms such as Random Forest, XGBoost, and logistic regression to assess the performance of these models in identifying similarities among pairs of questions. Secondly, we introduce a Siamese Long Short-Term Memory (LSTM) network to determine semantic similarity between current and archived questions, demonstrating its superior performance compared to previous methods.

TC07

Summit - 327

Innovative Approaches in Decision Making and Optimization

Contributed Session

Chair: Arik Sadeh, HIT Holon Institute of Technology, Holon, Israel

1 - Do People Underweight the Supply Line? An Eye Tracking Investigation

Heetak Shah, Indian Insititute of Technology Bombay, Mumbai, India, Tarikere T. Niranjan, Enno Siemsen

This study investigates possible *rational* reasons behind supply line "underweighting", well-known as a cognitive bias in supply chain decision-making. We believe that much of the notoriety surrounding this "bias" is perhaps misplaced, at least in contemporary supply chain decision making. We employ an eye tracking based experiment to understand how people consider the supply line. The main goal of this study is to understand if people fail to recognize the supply line or if they recognize it but deliberately disregard the information. This experiment further aims to understand if people disregard the supply line commensurate with the reliability or timeliness of the supply line. We also look at a possible solution to dampen the underweighting behavior in case of an untimely or unreliable supply line. This study's results will help increase our understanding of this underweighting behavior. The findings will also provide useful directions for future researchers to build on the vast prior literature in this domain and harness the insights for a better understanding of this important application area.

2 - Optimizing Inspection Strategies in Genetic Manufacturing Systems with Parallel Synthesis Lines Using Markov Decision Process

Mohammad Maydanchi, Auburn University, Auburn, AL, United States, Gregory Purdy, Daniel F. Silva

In recent years, synthetic biology has advanced significantly, with genetic engineering becoming an increasingly vital component of biotechnology. Creating an efficient production process for genetic constructs is crucial for the synthetic biology industry. Genetic Manufacturing Systems are a recent innovation that aim to streamline this process. Similar to other manufacturing processes, inspections are necessary for Genetic Manufacturing Systems to ensure high-quality outputs. This study uses a Markov decision process to model a Genetic Manufacturing System and determine the appropriate inspections for DNA fabrication. The model represents a multi-stage inspection allocation problem and examines how modifying a Genetic Manufacturing System's characteristics impacts the optimal inspection strategies. Additionally, we examine a Genetic Manufacturing System that creates two fragments simultaneously in two parallel synthesis lines and combines them to produce a longer fragment assembled based on the Gibson assembly protocol. The mathematical model presented provides a valuable tool for designing an optimal inspection strategy for Genetic Manufacturing Systems. This research has significant implications for professionals in the synthetic biology industry, as the findings can be used to optimize the production of high-quality genetic constructs.

3 - Network-driven airline on-time performance prediction with dynamic delay propagation

Yi Tan, The University of Alabama in Huntsville, Huntsville, AL, United States, Yajun Lu, Lu Wang

We study the impact of network information of recent delayed flights for airline's on-time performance prediction. As a critical performance indicator of air transport systems, flight delays have attracted great attention to the OM researchers. However, the impact of recent delays of other flights within the entire airport network remains largely unexplored. By addressing multiple challenges in airline operations, we proposed a novel network-based approach, which can derive a time-dependent delay risk score for each airport. Empirical experiments show that integrating this airline network extracted feature with a group of intrinsic features from the airline operation literature can significantly improve the flight delay detection ability. Also, SHAP analysis shows that the network extracted feature ranked as the most important among 13 benchmark predictors.

4 - An analysis on evolution rate of legal case citations in India

Kumar Sarthak, Indian Institute of Technology Delhi, New Delhi, India, Shaurya Shriyam, Nomesh Bolia

Ever changing citation behaviour of legal case citations results in a dynamic characteristic in the analysis of the citation network. The notion of landmark cases or important cases evolves with time according to the cases being cited as precedents in the judgment of the present case. Prior research has mainly focused on a statistical analysis of the network by applying traditional network metrics to quantify the importance of cases. Most statistical concepts and formalisms fail to capture the temporal evolution of the network. In this paper, we investigate the rate of evolution of landmark cases based on the time-varying graphs (TVGs) spanning successive time intervals. Additionally, we explore the structural properties of legal cases to capture the definition of importance by employing centrality measures such as relative degree, betweenness and longevity. Moreover, we evaluate the modularity of the network to detect the community structures in the network. For this study, we curate a large dataset of judgments from the Supreme Court of India on Article 300A (right to property) of the Indian Constitution. Our analysis shows that the temporal analysis, in association with the analysis of structural properties, provides a deeper understanding of the nuanced dynamics within the real-world dataset. Our work also identifies how the complexity of the citation network is directly proportional to complexity in legal cases due to multiple and hierarchical citations in cases and its evolution over time.

5 - Empirical and Theoretical Results on Bertrand Duopoly with Interaction

Arik Sadeh, HIT Holon Institute of Technology, Holon, Israel

The Bertrand competition is an arena where firms compete in the price market. In Cournout games, firms compete by setting their quantities in the market. The attractiveness of Bertrand's games in economics is the visibility of prices to all competitors.

Games are commonly used in markets with a few decision makers where the playing firms see downward demand functions.

In this study, we analyze two firms in a duopoly market. The firms sell nonidentical products that satisfy the same customers' needs.

The demand function for each firm is based on its price and the competitor's price, along with an interaction of both prices on each demand. We show that there is more than one Nash equilibrium point that leads to stability questions, The role of the cost function, its functional form, and parameters impact the number of potential equilibrium points. This is an essential issue in markets with technological and innovative products.

The dynamics of Bertrand's games can serve as a mechanism to reveal the costs of the competition.

TC08

Summit - 328

Cybersecurity Impacts and Strategies

Contributed Session

Chair: Madison Evans, Auburn University, Opelika, United States

1 - Demonstrating Cyber-Physical Attacks in a Multi-Fidelity Manufacturing Digital Twin Environment

Madison Evans, Auburn University, Auburn, AL, United States, Gregory Purdy

Cybersecurity has grown increasingly crucial due to the number of reported cyber incidents that occur every year. Every industry is vulnerable to the threats posed by cybercriminals. Some industries, like manufacturing, face additional challenges due to the prevalence of cyber-physical systems. These systems have components in both the physical and digital worlds, which create new and dangerous vulnerabilities for cybercriminals to exploit. Studies indicate that conducting security testing on these types of machines can lead to significant and costly damage. Manufacturers need a regulated and secure environment that closely emulates the physical properties of the system to investigate cyber-physical attacks and defense mechanisms. A digital twin of the system can be used to create such an environment; however, it introduces the question of the extent of modeling fidelity needed to secure systems. In this presentation, we will demonstrate a cyber-physical attack in a Digital Twin Environment with varying levels of fidelity. We will examine and differentiate the conclusions that can be made from the results at these different levels. We will use these results to construct a Multi-Fidelity Digital Twin that can be incorporated into a Cyber-Physical Manufacturing Range.

2 - The Impact of Security-Related Stress (Srs) on ICT-Enabled Productivity from The Transactional Model of Stress and Coping and Holistic Stress Model Perspective

Roya Ashrafidehkordi, University of North Texas, Denton, TX, United States

Information security demands within the organizations could lead to security-related stress (SRS), one of the critical issues in the information systems field in the past years. Previous research mostly worked on how this kind of stress would lead to employees violating the security policies and focused on the dark side. Given that, the present study will explore the influence of the SRS on ICT-enabled productivity to figure out the bright side of coping with SRS. The study integrates two theoretical frameworks, the transactional model of stress and coping and the holistic stress model, to examine the employees' appraisal mechanism, psychological perception, coping mechanism, and finally, the impact of this whole path on ICT-enabled productivity. This research would also apply different strategies for coping processes that would offer a comprehensive understanding of the multifaceted nature of stress and its impact on individual well-being and productivity. The research contributes to the existing literature by digging into the bright side of security-related stress and addressing the complex interrelationship between security-related stress and ICT-enabled productivity. Overall, the article comprehensively explores various coping mechanisms to advance understanding of this critical intersection.

3 - Does the CEO regulatory focus matter in containing data breaches? Evidence from US-listed firms

Amit Kumar, Indian Institute of Management Ahmedabad, Ahmedabad, India, Adrija Majumdar, Indranil Bose

In the current times, data breaches are pervasive, affecting many organizations operating in diverse sectors. Data breaches cause significant damages, including financial setbacks, reputation loss, erosion of customer trust, and legal liabilities. Surprisingly, top management still treats data breaches and cybersecurity as a technical function rather than a strategic one. This misconception and the enormous costs of containing a data breach underscores the need for CEOs to lead cybersecurity efforts. Data breach literature has neglected the impact of CEOs on data breaches, and our research tries to fill this gap. Drawing on regulatory focus theory, we examine how CEO regulatory focus- promotion and prevention- impacts data breach risk. We contextualize our study by investigating the moderating role of IT investment in this relationship. Using the data of US-listed firms over six years (2016-2021), we find that CEO promotion focus reduces data breach risk, while CEO prevention focus increases the data breach risk. We also find that IT investment mitigates the impact of both the CEO promotion focus and the CEO prevention focus on data breach risk. By emphasizing the containment of data breach risk as a firm performance metric, our study highlights the importance of cybersecurity leadership at the top management level. We contribute to the literature on data breach and regulatory focus by integrating CEO motivation (measured through regulatory focus) and IT investment with data breach dynamics while providing insights for practitioners and policymakers in cybersecurity strategies.

4 - The Impact of Information Security Investment on Firm Financial Risk: The Role of Board Diversity

Niyanta Sookraj, Auburn University, Auburn, AL, United States, Yen-Yao Wang, Xin Luo

As information security threats rampage through the digital era, their impact on firm financial risk has become increasingly concerning. Despite the considerable surge in information security investments, data breaches have continued to persist alongside other security concerns, raising doubts about the efficacy of these investments. This study examines the relationship between information security investments and firm financial risk through the lens of corporate governance in the form of board diversity. The value of board diversity lies in its potential to bring about a broader range of perspectives, expertise, and innovation through the facilitation of discussions and the exchange of ideas, thus strengthening the board's problem-solving and decision-making abilities. Specifically, we focus on one element of cognitive diversity – board tenure diversity. To the best of our knowledge, this is the first study to examine board tenure diversity in the context of firm financial risk. Utilizing a ten-year panel encompassing 7,892 firm-year observations across 1,071 unique firms, we uncover that information security investments lower financial risk for firms. Furthermore, we find that board tenure diversity is a significant moderator in the relationship between information security investments and firm financial risk, indicating that the board's composition can play a strategic role in risk management. Our results suggest that firms should prioritize information security investments and consider board composition to enhance the effectiveness of these investments. By incorporating board tenure diversity, firms can more effectively navigate the complex landscape of information security threats.

5 - Filling the Blanks: Enhancing Vulnerability Assessment in Dark Hosts

Mohammad Abdullah, The Ohio State University, Dublin, OH, United States, Theodore Allen, John McCarty

Our research addresses significant vulnerabilities in cybersecurity assessments, focusing on 'dark hosts'—network hosts that evade regular scans. Analyzing a dataset from a major Midwest institution featuring over 48,495 hosts, we found that more than 50% of the hosts went unscanned in any given month. To address these gaps, we deployed machine learning models that effectively predict potential vulnerabilities using historical scan data. Our classification models successfully identified critical vulnerabilities with over 91% accuracy. By leveraging information such as the host's operating system, past scan data, and organizational context, these models significantly enhance the predictability of vulnerabilities in dark hosts. This not only improves our ability to secure network infrastructures but also provides crucial insights to help cybersecurity management teams in any organization make better-informed decisions regarding their next steps in vulnerability management and prevention.

TC09

Summit - 329

Robust and Ethical Market Design

Invited Session

Auctions and Market Design

Chair: Juba Ziani, Georgia Tech, Atlanta, GA, United States

Co-Chair: Rachel Cummings, Columbia University, New York, NY, United States

1 - Equilibria of Data Marketplaces with Privacy-Aware Sellers Under Endogenous Privacy Costs

Diptangshu Sen, Georgia Institute of Technology, Atlanta, GA, United States, Jingyan Wang, Juba Ziani

We study a two-sided data ecosystem comprised of an online platform, users who may want to participate on the platform, and downstream data buyers who look to purchase data from the platform. The platform incentivizes users to participate by providing them with a service. It also acts as an aggregator, collecting data from participating users and selling them downstream. Therefore, potential users face the following trade-off - do they abstain from joining and miss out on the utility provided by the platform, or incur privacy costs by participating and sharing their data ?

In light of the rapidly changing privacy attitudes of users towards online platforms, we introduce a novel modeling element for such two-sided markets: the privacy costs of participating users in our model are ***endogenous*** and depend on ***how many downstream buyers gain access to their data.*** Our main contributions are the following: i) providing a complete characterization of marketplace equilibria for two classes of user utility functions (when each user gets a fixed utility from participating, and when the utility is linearly increasing in the number of other users that participate) and ii) providing a comparative analysis with the setting where privacy costs are exogenous and fixed. We demonstrate that market equilibria change significantly when endogeneity in user privacy costs are accounted for. Finally, we conclude by providing simulations and semi-synthetic experiments that extend our results to more general assumptions. In particular, we experiment with different distributions of user privacy costs and different functional forms of user utilities.

2 - Repeated Contracting with Multiple Non-Myopic Agents: Policy Regret and Limited Liability

Varun Gupta, University of Pennsylvania, Philadelphia, PA, United States, Natalie Collina, Aaron Roth

We study a repeated contracting setting in which a Principal adaptively chooses amongst k Agents at each of T rounds. The Agents are non-myopic, and so a mechanism for the Principal induces a T-round extensive form game amongst the Agents. We give several results aimed at understanding an under-explored aspect of contract theory -- the game induced when choosing an Agent to contract with. First, we show that this game admits a pure-strategy \emph{non-responsive} equilibrium amongst the Agents -- informally an equilibrium in which the Agent's actions depend on the history of realized states of nature, but not on the history of other's actions. Next, we show that if the Principal selects Agents using a \emph{monotone} bandit algorithm, then for any concave contract, in any such equilibrium, the Principal obtains no regret to contracting with the best Agent in hindsight -- not just given their realized actions, but also to the counterfactual world in which they had offered a guaranteed T-round contract to the best Agent in hindsight, which would have induced a different sequence of actions. Finally, we show that if the Principal selects Agents using a monotone bandit algorithm which guarantees no swap-regret, then the Principal can additionally offer only limited liability contracts (in which the Agent never needs to pay the Principal) while getting no-regret to the counterfactual world in which she offered a linear contract to the best Agent in hindsight -- despite the fact that linear contracts are not limited liability.

3 - Bayesian Strategic Classification

Saeed Sharifi-Malvajerdi, Toyota Technological Institute at Chicago, Chicago, IL, United States, Lee Cohen, Kevin Stangl, Ali Vakilian, Juba Ziani

In strategic classification, agents modify their features, at a cost, to obtain a positive classification outcome from the learner’s classifier, typically assuming agents have full knowledge of the deployed classifier. In contrast, we consider a Bayesian setting where agents have a common distributional prior on the classifier being used and agents manipulate their features to maximize their expected utility according to this prior.

The learner can reveal truthful, yet not necessarily complete, information about the classifier to the agents, aiming to release just enough information to shape the agents' behavior and thus maximize accuracy. We show that partial information release can counter-intuitively benefit the learner's accuracy, allowing qualified agents to pass the classifier while preventing unqualified agents from doing so.

Despite the intractability of computing the best response of an agent in the general case, we provide oracle-efficient algorithms for scenarios where the learner's hypothesis class consists of low-dimensional linear classifiers or when the agents' cost function satisfies a sub-modularity condition. Additionally, we address the learner's optimization problem, offering both positive and negative results on determining the optimal information release to maximize expected accuracy, particularly in settings where an agent's qualification can be represented by a real-valued number.

4 - The Hidden Costs of Privacy Choice for Marginalized Groups

Tamalika Mukherjee, Columbia University, New York, NY, United States, Rachel Cummings, Talia Gillis

In both the U.S. and Europe, efforts to regulate privacy concerns often rely on a `notice and consent' approach. Under this approach, organizations are required to provide clear and understandable notices to users about their data practices, including what information is being collected, how it is being used, and with whom it is being shared. Users are then given the opportunity to consent to these practices. While there is a rich literature that studies the potential harms caused by individuals giving consent (in the context of privacy notices) and the role of consent itself in shaping privacy policies, little attention has been paid to the negative impact of choice on marginalized groups, i.e., the consequences of individuals from marginalized groups withholding consent. When the collection and usage of information provides for some group benefits, the individual's decision to withhold information can create information externalities. We consider situations in which the negative externalities resulting from the withholding of personal information disproportionately impact marginalized groups,and develop a mathematical framework that encapsulates these harms.

5 - You Can Have Your Cake and Redistrict It too

Jamie Tucker-Foltz, Harvard University, Allston, MA, United States, Gerdus Benade, Ariel Procaccia

The design of algorithms for political redistricting generally takes one of two approaches: optimize an objective such as compactness or, drawing on fair division, construct a protocol whose outcomes guarantee partisan fairness. We aim to have the best of both worlds by optimizing an objective subject to a binary fairness constraint. As the fairness constraint we adopt the geometric target, requiring the number of seats won by each party to be at least the average of its outcomes under its worst and best possible partitions of the state.

To study the feasibility of this approach, we introduce a new model of redistricting, closely mirroring the classic model of cake-cutting. This model has two innovative features. First, in any part of the state there is an underlying "density" of voters with political leanings toward any given party, making it impossible to finely separate voters for different parties into different districts. Second, parties may disagree on the distribution of voters. In the absence of a "ground truth" distribution, a redistricting algorithm must therefore aim to simultaneously be fair to each party with respect to its own reported data. Our main theoretical result is that, surprisingly, the geometric target is always feasible with respect to arbitrarily diverging datasets on how voters are distributed.

Our empirical results, which use real election data and maps of six US states, demonstrate that the geometric target is always feasible, and that imposing it as a fairness constraint comes at almost no cost to three well-studied optimization objectives.

TC10

Summit - 330

Enhancing Health Outcomes and Performance Through Integration and Analysis

Contributed Session

Chair: Yoon Sang Lee, Columbus State University, Columbus, GA, United States

1 - Controlled Under-Sampling with Stacking Ensemble Learning for Class Imbalance Problem

Yoon Sang Lee, Columbus State University, Columbus, GA, United States, Riyaz Sikora

Imbalanced data sets are a growing problem in data mining and business analytics. However, their ability to predict the minority class deteriorates in the presence of class imbalance. Although there have been many approaches that have been studied in literature to tackle the imbalance problem, most of these approaches have been met with limited success. In this study, we propose three methods based on a wrapper approach by combining the use of under-sampling with ensemble learning to improve the performance of standard data mining algorithms. We test our ensemble methods on 10 data sets collected from the UCI repository with an imbalance ratio of at least 70%. We compare their performance with two other traditional techniques for dealing with the imbalance problem and show significant improvement in the recall, AUROC, and the average of precision and recall.

2 - Wealth Shocks and Physician Behavior: Evidence from Childbirth

Yanhao Wang, Indiana University Bloomington, Bloomington, IN, United States

Little is known about how physicians' personal wealth affects their behavior. Using data on physicians' housing returns linked to their treatment choices, this paper provides new empirical evidence on the sensitivity of physicians' behavior to their wealth and studies the policy implications on medical expenditure and patient health. Specifically, I look at obstetricians who have large discretion and significant financial incentives in choosing cesarean sections over vaginal deliveries. For identification, I rely on the variation in obstetricians' housing returns arising from the Great Recession, depending on when or where they purchased their houses ex-ante. Using a patient-level regression model with obstetrician fixed effects, I estimate that a one-standard-deviation decrease in obstetricians' housing wealth leads to an increase of 1.8 percentage points, or 4.5%, in C-section probability. The increased C-section rate leads to longer in-hospital stays and higher infection rates among the patients. Finally, I show in two counterfactuals that the estimated wealth sensitivity is useful in predicting the effectiveness of price regulations and alternative payment models.

3 - Racial Disparities and Comorbidities: Network Analysis of Maternal Outcomes in Alabama

Yasin Fatemi, Auburn University, Auburn, AL, United States, Haneen Ali, Jingyi Zheng

Introduction The study employed a robust network analysis methodology to assess the effects of race and comorbidities on birth outcomes, using a dataset of 443,902 mothers in Alabama from 2014 to 2021.

Methods Four multimorbidity networks corresponding to White, Black American, Asian, and American Indian or Alaskan Native groups were constructed to explore distinct comorbidity patterns. The nodes in these networks represented various diseases, while the edges, quantified by the Salton Cosine Index, depicted the associations between these conditions. Additionally, two separate networks were analyzed for low birth weight (LBW) and normal birth weight (NBW) to identify the differential impacts of specific diseases. Feature selection methods including Random Forest and Logistic Regression were applied to pinpoint crucial intersections between the LBW and NBW networks, enhancing the granularity of the analysis.

Results The findings indicated significant racial disparities in the density of comorbidity networks, with more complex disease interactions observed among Black, Indian, and Asian groups compared to Whites. Preexisting hypertension and eclampsia emerged as significant risk factors for LBW in White and Black groups, while gestational hypertension was prevalent across multiple racial groups. The LBW network displayed greater density than the NBW network, highlighting the intricate connections between comorbidities leading to adverse birth outcomes.

Conclusion These insights underline the necessity for healthcare interventions tailored to the distinct health profiles of each racial group to effectively address and reduce maternal health disparities.

4 - Assocation between Medicaid policy regarding dental benefits and tooth retention

Marta Ventura, Penn State University, University Park, PA, United States, Paul Griffin

Extensive tooth loss affects function, general health, and quality of life. The World Health Organization considers retaining 20 natural teeth as necessary for functional dentition. Increasing access to clinical dental care among adults could significantly reduce tooth loss. The two most common diseases that cause tooth loss, such as dental caries and periodontal disease. We use a binomial ordinal logistic regression to examine the association between Medicaid policy regarding dental benefits and tooth retention. For this model, the exposure variable is based on whether person had continuous Medicaid exposure from 2007 to 2012 and from 2013 to 2018. The data that we used came from the Centers for Disease Control and Prevention (CDC) survey called Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS includes one question on the number of teeth lost due to tooth decay or gum disease. We used the response to this question to develop two categories of tooth retention: functional dentition (0 to 5 missing teeth) and not having functional dentition (6 or more missing teeth). This model accounts for change in 'Comprehensive Benefits' over time; and uses high-income (Non-Medicaid) as a control. The study population are adults 35 years old - 54 years old. The 'exposure' is 1) no benefits in *either* period 2007-2012 and 2013- 2018; 2) continuous benefits in *both* periods 2007-2012 and 2013- 2018; 3) *no* benefits in 2007-2012 and *benefits* in 2013-2018 (meaning they *gained* benefits); and 4) *benefits* in 2007-2012 and *no* benefits in 2013-2018 (meaning they *lost* benefits).

TC11

Summit - 331

Operational Planning of Emerging Vehicle Technologies

Contributed Session

Chair: Heeseok Moon, Kyunghee University

1 - Solving optimal relocation transport problem in shared mobility services

JOON MOON, The Ohio state university, Columbus, OH, United States, Qadeer Ahmed

The rise of micromobility sharing services such as scooters and bike sharing, exemplified by companies like Lime and Uber, has led to significant global demand for shared mobility. However, these services often struggle with supply-demand imbalances, particularly in locations and times where there is an excess of supply or demand. Therefore, relocating vehicles, often conducted overnight, is crucial to ensure optimal service coverage. This paper delves into the problem of vehicle relocation in shared mobility services to minimize transport costs. To solve this problem, this study proposes a problem formulation and approach that is grounded in the optimal transport theory. When comparing the conventional assignment problem formulation, which treats each demand and supply as separate entities and can be deterministic and computationally burdensome, the optimal transport formulation is expected to simplify large-scale problems while maintaining its utility. We highlight the effectiveness of our proposed approach in comparison to conventional methods based on the mathematical formulation of assignment problems. Additionally, these methodologies are validated using data from a bike-sharing service in Seoul, South Korea, providing practical insights into effective relocation strategies for addressing clustered demand and supply challenges.

2 - An enhanced adaptive large neighborhood search procedure for a rich vehicle routing problem

Gizem Ozbaygin, Sabanci University, Istanbul, Turkey, Saeedeh Ahmadi Basir

Motivated by a real-life application, this study focuses on a rich vehicle routing problem characterized by a multi-period planning horizon, time window constraints, visual attractiveness and driver consistency concerns. We propose an enhanced adaptive large neighborhood search procedure to tackle the problem and report our findings based on computational experiments.

3 - Decision Support Framework in Electric Vehicle Charging Technologies Using Patent Data: A Multi-criteria Approach Considering National Technology Portfolio, Technology Impact and Development Stage

Heeseok Moon, Kyunghee University, Seoul, Korea, Republic of, Hyunhong Choi

In the context of global efforts towards carbon neutrality and climate change response, the strategic importance of charging technologies for expanding electric vehicle (EV) deployment is increasingly recognized. This study identifies 40 potential technology topics using Structural Topic Model (STM) on 10,393 patent data related to EV charging technology over 12 years (2010-2021) collected by the USPTO. The correlation between these topics reveals seven clusters with distinct technological properties. Based on this, a comprehensive comparison of the technology portfolios of six major countries (US, Japan, France, Germany, Korea, China) was conducted, highlighting each country's strengths and strategic focus areas. Emerging technologies were identified through patent application year and citation data, providing an overview of the EV charging technology landscape.

To develop a comprehensive decision-making framework, technologies were classified into embryonic, growth, and maturity stages through a technology life cycle assessment. This framework utilized country-specific technology focus and topic influence to present suitable strategies for each country. This study not only provides a comprehensive overview of global technology development trends but also serves as an essential tool for investment support and policy development, aiding in the establishment of customized industrial strategies. Furthermore, it supports the sustainable expansion of EV charging infrastructure.

This research makes three key contributions: it conducts the first patent analysis of EV charging technologies, provides visualized results for policymakers through technology cluster analysis and national technology portfolio comparisons, and proposes an effective screening tool for national R&D investment by introducing the concept of the technology life cycle

TC12

Summit - 332

Intermodal and Sustainable Freight Logistics

Contributed Session

Chair: Hyunseop Uhm, Argonne National Laboratory, Lemont, IL, United States

1 - Intermodal Inbound Freight Allocation Optimization for Port-Hinterland Distribution Planning

Sarita Rattanakunuprakarn, University of Tennessee, Knoxville, Knoxville, TN, United States, Mingzhou Jin

Improving freight distribution between ports and hinterlands is essential for optimizing transportation networks. This research introduces an innovative approach for optimizing inbound freight allocation and facility location selection within the port-hinterland distribution framework of the U.S. Southwest region, specifically emphasizing the network originating from the San Pedro Bay port complex. By utilizing a Stochastic Mixed Integer Programming (MIP) model, our approach allocates goods for transportation via railroads and trucks, while also considering the availability of distribution centers. Unlike traditional practice that primarily rely on port proximity, our proposed network integrates hinterland distribution centers into the logistical framework, thereby refining container classification and unit-train formation processes. The model evaluates performance based on transportation efficiency and cost-effectiveness, taking into account economic, environmental, and social factors. This study contributes to the advancement of freight distribution strategies, thereby enhancing the resilience and sustainability of port-hinterland logistics networks.

2 - Sustainable collaborative urban supply chains (STRAUSS): A comparative study for waste collection

Dilay Aktas, KU Leuven Institute for Mobility, Leuven, Belgium, Pieter Vansteenwegen

In the STRAUSS project, we develop mathematical models and algorithms to optimize urban logistics operations. Unlike existing studies, STRAUSS specifically targets the urban context and considers the measures implemented by city administrations that impact logistics operations within their territory. These measures may include car-free zones, designated parking spaces or time windows for truck deliveries, and bans on certain vehicles in specific areas. Our algorithms evaluate the impact of these measures to mitigate the negative effects of logistics operations on city livability and enhance the competitiveness of supply chain actors. Contrary to existing urban logistics optimization problems, our models are grounded in an in-depth economic analysis. This work focuses on one crucial stream of daily logistics operations: waste collection. Typically, waste collection trucks enter the city empty to collect waste, while delivery trucks from supply chain actors return empty to their depots. Consequently, trucks are on average only half-full during their transportation routes. We explore the integration of deliveries to retail stores in the city center with waste collection, examining how this scheme could be practically implemented and identifying potential drawbacks. We develop a routing-optimization model to generate combined collection-and-delivery routes and evaluate these scenarios using a case study.

3 - Advanced Multimodal Urban Transport: Synchronizing Freight Delivery and Passenger Movement

Nastaran Tork, University of Minnesota, Minneapolis, MN, United States, Chengwenjian Wang, Alireza Khani, Jean-Philippe Richard

Urban congestion and pollution are critical challenges demanding innovative, interconnected transportation solutions. This study proposes an integration of freight delivery into the light-rail transit network, optimizing the system to accommodate both passengers and parcels. The approach involves the use of strategically placed lockers at train stations where parcels can be dropped off by first-mile services or customers and then delivered to destination lockers via trains. The core objective is to create a decision-support framework that adaptively distributes and routes parcels and minimize total delivery time, costs and number of selected stations and crowd-shipping services adapted for freight transportation. This framework incorporates various uncertainties, including fluctuating parcel volumes, specific delivery time requirements, and last-mile logistics constraints. It also considers the dynamic availability of train capacity, considering both current and predicted passenger numbers. These variables are modeled through a range of scenarios to ensure flexibility. We propose a stochastic optimization model for dynamic routing, enhanced by predictive analytics for accurate demand forecasting. We discuss the formulation of the model, including the key constraints and objectives, as well as the methods used to solve it. Computational results on a case study from a major city are presented to demonstrate the practical applicability and benefits of the approach.

4 - Implementation and Analysis of Urban Freight Electrification for the E-commerce Delivery using POLARIS Transportation Framework

Hyunseop Uhm, Argonne National Laboratory, Lemont, IL, United States, Abdelrahman Ismael, Natalia Zuniga-Garcia, Olcay Sahin, James Cook, Joshua Auld, Monique Stinson

Freight truck electrification for last-mile delivery is one of the most important research topics to reduce the dependency on fossil fuel operations. Although a battery electric truck still has limitations on daily operations with lower driving ranges and higher purchasing cost than a conventional truck, operations with electrified trucks reduce total energy usage and recharging/refueling cost. In this paper, we develop and implement a fleet-mix and multi-shift e-commerce delivery model into POLARIS, an agent-based transportation modeling framework. Based on the e-commerce delivery orders synthesized by POLARIS demand model, a fleet owner decides the best electrification ratio of the fleet and the proper number of chargers which gives the minimum total cost. The initial truck operation planning is described as mixed-integer linear programming model including common constraints for heterogeneous vehicle routing problem, multi-shift operation with recharging scheduling, and recharger allocation. A bi-level VNS-TS heuristic with the event trigging and memory managing features in POLARIS is also suggested for efficient solution search. All truck tours for daily e-commerce delivery are simulated onto the POLARIS transportation network with other passenger travel, energy use and battery level of every truck is tracked at every simulation timestep, and route replanning and en-route recharging models are added to modify the initial plan for the route feasibility. Finally, the best EV ratio with different EV driving ranges, recharging/refueling cost, and public recharger availability scenarios are evaluated in multiple POLARIS metropolitan cities.

TC13

Summit - 333

Circular Economy and Sustainable Supply Chains

Contributed Session

Chair: Geng Deng, Harbin Engineering University, Harbin, N/A

1 - Managerial Efficiency and Ordering Policies in a Circular Economy

Sergey Naumov, Smeal College of Business, The Pennsylvania State University, University Park, PA, United States, Saurabh Bansal, V Daniel R Guide

A Circular Economy (CE) is a system where products are collected after customers have used them, remanufactured, and then put back in circulation for further use. This reduces the need for virgin materials to manufacture new units and wastage from used units. However, it is well-known that delays in inventory returns often lead managers to make suboptimal ordering decisions. This problem is important in CE systems, where poor inventory management can reduce or completely negate CE benefits, but existing literature lacks evidence on its magnitude and key contributing factors. Addressing this gap is essential for developing operational details and maximizing benefits of CE systems. Using a novel complexity ordering, we investigate the quality of managerial decision-making. In our experimental study, participants made inventory replenishment decisions in various CE systems that differed in the complexity of reverse-flow of goods. We find that participants' operating costs were three to five times higher than normative costs, and, as CE system became more complex in reverse-flow, the suboptimality of managerial decision-making worsened. Our experimental results provide prescriptions and guidance for practitioners by identifying: (i) CE configurations for which managerial decision-making tends to be better, and (ii) behavioral nudges to improve the quality of decision-making.

2 - Location-Routing Strategies for Drug Take-Back Programs in Rural Area

Jihyun Jo, National Institute of Standards and Technology, Gaithersburg, MD, United States, Evan Wallace

The circular economy is a sustainable model that emphasizes sharing, leasing, reusing, repairing, refurbishing, and recycling existing materials and products. One of aims is protecting environment by minimizing waste. Despite the pharmaceutical industry's significant contribution to global health, it is a major source of river water pollution. Pharmaceutical waste, including drugs and metabolites, enters river through improper disposal practices. Waste is generated throughout the pharmaceutical supply chain due to factors such as overproduction, excess stocks, improper storages, and overprescription, etc. In the United States, drug take-back program and state prescription drug repository programs are implemented, and these programs reduce the pharmaceutical waste. However, they face challenges, including the availability of drop-off locations, which vary by region and are often limited in rural areas compared to urban areas. To address this challenge, we propose a location-routing model specifically tailored for rural areas. We consider two heuristics: one that prioritizes determining drop-off locations before planning individual routes using the multi-depot vehicle routing problem, and another that plans individual routes first and then identifies suitable drop-off locations. Throughout the location-finding process for the second heuristic, we continuously optimize individual routes to enhance the overall solution.

3 - Does Corporate Cooperation Culture Foster Supply Chain Resilience? A Large-Scale Empirical Study

Deng Geng, Harbin Engineering University, Harbin, China, People's Republic of, Liu Weiwei, Lee Ruby

Despite firms' efforts to maintain their supply chains in complex business environments, they are now more vulnerable to supply chain disruptions than ever before and have begun to pay attention to how to enhance resilience against such disruptions. The existing literature has limited attention on how corporate cooperation culture affects supply chain resilience, and large-scale empirical evidence capturing the complex relationship between corporate cooperation culture and supply chain resilience remains elusive. Our research begins to bridge this gap by adopting a textual analysis approach to construct metrics for measuring corporate cooperation culture and examining its interplay with

supply chain resilience. Based on empirical testing of data from 4751 Chinese listed companies, we found that corporate cooperation culture is positively correlated with supply chain resilience. Further research revealed that supply chain transparency plays a mediating role, and corporate ESG disclosure partially moderates the impact on supply chain resilience. These findings significantly expand the literature on supply chain resilience and contribute to a better understanding of the comprehensive impact of corporate cooperation culture on supply chain resilience.

TC14

Summit - 334

Financial Decision Making and Risk Management

Invited Session

Finance

Chair: Dohyun Ahn, CUHK, Shatin, Hong Kong

1 - Dynamic Pricing in Equity Crowdfunding: Optimizing Success in Blockchain Ventures

Bretislav Hajek, National University of Singapore, Singapore, Singapore, Jussi Keppo, Steven Kou

Equity crowdfunding has become a favored method for blockchain-based companies to raise initial capital by offering governance tokens as equity. This model enables real-time price and marketing adjustments in response to demand, bypassing the need for an investment bank typically required for an IPO. Our research models equity crowdfunding as a continuous-time problem, examining its optimal pricing strategy. We demonstrate that dynamic pricing increases fundraising effectiveness and the success rate of crowdfunding campaigns. Furthermore, we assess the moral hazard associated with this approach and investigate the potential advantages for investors.

2 - (Almost) Model-Free Dynamic Mean Quadratic Variation Analysis of Log Returns

Chihoon Lee, Stevens Institute of Technology, Hoboken, NJ, United States, Zhenyu Cui, Dongwoo Kim, Yanchu Liu

In this paper, we propose an almost model free dynamic mean quadratic variation (MQV) asset allocation analysis for log returns, which is named log-MQV. It has several advantages, including recovering the time-consistency in the optimal investment decision, conforming to investment wisdom, and the explicit closed-form optimal investment decisions for most stochastic models in finance under both complete and incomplete market settings. The proposed model covers multivariate Ito diffusion models, and can further incorporate jump risks, regime switching and stochastic volatility features through a unified framework. We also illustrate that the proposed framework allows for a fully data-driven implementation utilizing historical time series data, and this paves the path for a fully data-driven robo-advising investment strategy. Extensive numerical and empirical experiments illustrate the performance of the proposed optimal log-MQV portfolio.

3 - Robust Optimal Strategies for Two-Period Liquidation in Financial Systems

Hongyi Jiang, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, Dohyun Ahn

We study the problem of two-period liquidation strategies in financial systems. During financial crises, asset liquidation is often inevitable but can lead to substantial losses if a significant amount of assets are sold simultaneously at depressed prices. To tackle this challenge, we consider a two-period liquidation model that allows for preemptive liquidation before maturity and propose a worst-case approach to the associated liquidation problem. Specifically, we propose a tractable liquidation strategy that maximizes a bank's worst-case cash position after liquidation, taking into account the uncertainty of other banks' liquidation decisions. In addition, we find that the unique Nash equilibrium is attained when all banks adopt this proposed strategy. We further demonstrate that the worst-case optimal strategy retains a similar form even when interbank liabilities are involved, where we consider both scenarios of full network information and partial network information. While our analysis builds upon a stylized model, our findings offer valuable guidelines for developing robust liquidation strategies that mitigate losses resulting from asset liquidation.

4 - Stressed Distribution Selection via Causal Optimal Transport

Eunji Kwon, Korea Advanced Institute of Science and Technology, Daejeon, Korea, Republic of, Kyoung-Kuk Kim

Recently, causal optimal transport has emerged as an alternative to classical optimal transport for capturing a temporal information structure between probability distributions within their Wasserstein distance. Utilizing this framework, we focus on identifying a stressed distribution with minimal transportation cost. We demonstrate primal attainment and present a dual reformulation incorporating entropic regularization, establishing the strong duality. This reformulation facilitates the computation of optimal solution to our problem.

TC15

Summit - 335

Data Science for Operations

Invited Session

Revenue Management and Pricing

Chair: Ruihao Zhu, Cornell University, Ithaca, NY, United States

1 - Pricing Analytics of Primary and Ancillary Products Using Conversion Rate Data

Changchun Liu, National University of Singapore, Singapre, Singapore, Hailong Sun, Chung Piaw Teo, Maoqi Liu

This paper focuses on a firm that sells both primary and ancillary products to customers, with the condition that customers can purchase the ancillary products only if they purchase the primary ones. Our goal is to jointly determine the prices for all the products to optimize profits to the firm. This problem is challenging as it involves customer choice over subsets (primary product + ancillary services) and precedence constraints on purchase decisions. To tackle these challenges, we use a class of additive perturbed utility model (APUM) to explicitly capture the complex customer choice scenarios. We show that the ground-truth choice model can be uniformly approximated by an APUM under suitable reformulation. To facilitate the model calibration, we construct a tractable APUM by introducing slack variables of the choice constraints. In particular, we leverage the separable structure of APUM and design a polynomial number of price experiments to obtain market share shift information, which allows us to calibrate the APUM model with shape constraints. By utilizing piecewise linear approximation, we demonstrate that the resulting data-driven pricing problem can be formulated as a mixed integer program, which can be efficiently solved by commercial solvers. Through extensive numerical experiments, we provide compelling evidence of the superior performance of our framework compared to other methods. This opens up a new way to solve the ancillary pricing problem (and potentially other multidimensional pricing problems under complex customer choice behavior) in a data-driven manner.

2 - Two-stage Online Reusable Resource Allocation: Advance Reservation, Overbooking and Confirmation Call

Ruicheng Ao, Massachusetts Institute of Technology, Cambridge, MA, United States, Hengyu Fu, David Simchi-Levi

We investigate an online reusable resource allocation problem with advance reservations, modeling it as a two-stage admission control problem. In Stage I, customer requests arrive randomly over time at a nonstationary rate to reserve one unit of capacity for a random duration. The number of accepted reservations may exceed the available capacity, a practice known as overbooking. In Stage II, both customers with reservations and walk-in customers arrive to check in sequentially. Reservations for day k can be canceled by customers without penalty before the confirmation call, which is made by the decision maker (DM) at some time before the day's service ends. The DM must make admissions decision upon receiving booking requests or check-ins. If a check-in from a reservation is rejected, an overbooking loss occurs. Our objective is to develop an online admission control policy that maximizes expected total revenue. We propose a dynamic thresholding policy that achieves constant regret compared to the optimal offline policy. Furthermore, we demonstrate that, with a logarithmically smaller scale of walk-in customers, no online policy can achieve better regret than Ω(T). We make a sensitivity analysis on the confirmation time, revealing that the DM can delay confirmation until a small logarithmic interval before the end of the day while still achieving near-optimal performance. Additionally, we conduct numerical experiments to examine the relationships between reservation and walk-in arrival rates, the confirmation time, and the performance of our proposed algorithm.

3 - Endogenous State Transitions in Transfer Learning and Dynamic Pricing

Shiyuan Wang, Tsinghua University, Beijing, China, People's Republic of, Jing-Sheng Song, Bora Keskin

This paper studies transfer learning in the context of dynamic pricing and endogenous state transitions. We consider a seller offering a large number of related products that are sold sequentially, each over a discrete time horizon. The expected demand for a product is a function of the product's price and a state variable that depends on past product prices. For each product, the parameters of the demand function are drawn from a common distribution based on the state variable. The seller does not know the underlying demand parameters and needs to learn them from accumulating noisy demand observations. The salient feature of our study is the endogenous state transitions: the state variable characterizes the key uncertainty in each product's pricing problem, and the state transition from one product to the next depends on previous pricing decisions. We develop an algorithm that simultaneously learns the common distribution and the demand parameters for each product to maximize cumulative revenue. We also establish an upper bound on the regret of our algorithm as well as a matching lower bound on regret, which implies that our algorithm is rate-optimal. Numerical experiments demonstrate our algorithm's good performance.

4 - Satisficing Exploration in Bandit Optimization

Qing Feng, Cornell Tech, New York, NY, United States, Tianyi Ma, Ruihao Zhu

Motivated by the concept of satisficing in decision-making, we consider the problem of satisficing exploration in bandit optimization. In this setting, the learner aims at finding a satisficing arm whose mean reward exceeds a certain threshold. The performance is measured by satisficing regret, which is the cumulative deficit of the chosen arm's mean reward compared to the threshold. We propose SELECT, a general algorithmic template for Satisficing Exploration via LowEr Confidence bound Testing, that attains constant satisficing regret for a wide varieties of bandit problems even when the arm space is large and possibly infinite (whenever a satisficing arm exists). Specifically, given a class of bandit optimization problems and a corresponding (standard) regret minimization oracle, SELECT iteratively makes use of the oracle to identify a potential satisficing arm. Then, it collects data samples from this arm, and continuously compares the lower confidence bound of the identified arm's mean reward against the threshold value to determine if it is a satisficing arm. As a complement, SELECT also enjoys the same (standard) regret guarantee as the oracle whenever no satisficing arm exists. Finally, we conduct numerical experiments to validate the performance of SELECT for several popular bandit optimization settings.

5 - Near-Optimal Policies for CMDP with Applications in RM

Jiashuo Jiang, Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of

Constrained Markov Decision Process has been a popular framework for online decision making. In this problem, the decision maker interacts with an unknown environment, where the state of the environment evolves in a Markovian manner. The goal of the decision maker is to maximize the total accumulated reward subject to some long-term constraints. It finds recent applications in modeling resource allocation problems. In this talk, I will present a recent advance in the study of the learning problem for CMDP. The main result is an instance-dependent $\tilde{O}(1/\epsilon)$ sample complexity for approximating the optimal policy. Several new elements are introduced to achieve this improved result, including i) a LP basis characterization of instance hardness, ii). a modified resolving heuristics with a stablized optimal basis. This is based on a joint work with Yinyu Ye.

TC16

Summit - 336

Advances in Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Omar Mouchtaki, New York University, New York, NY, United States

Co-Chair: Pnina Feldman, Darden School of Business, UVA, Charlottesville, VA, United States

1 - Interpolating Item and User Fairness In Multi-Sided Recommendations

Qinyi Chen, Massachusetts Institute of Technology, Cambridge, MA, United States, Jason Cheuk Nam Liang, Negin Golrezaei, Djallel Bouneffouf

Today's online platforms rely heavily on algorithmic recommendations to bolster user engagement and drive revenue. However, such algorithmic recommendations can impact diverse stakeholders involved, namely the platform, items (seller), and users (customers), each with their unique objectives. In such multi-sided platforms, finding an appropriate middle ground becomes a complex operational challenge. Motivated by this, we formulate a novel fair recommendation framework, called Problem (FAIR), that not only maximizes the platform's revenue, but also accommodates varying fairness considerations from the perspectives of items and users. Our framework's distinguishing trait lies in its flexibility: it allows the platform to specify any definitions of item/user fairness that are deemed appropriate, as well as decide the "price of fairness" it is willing to pay to ensure fairness for other stakeholders. We further examine Problem (FAIR) in a dynamic online setting, where the platform needs to learn user data and generate fair recommendations simultaneously in real time, which are two tasks that are often at odds. In face of this additional challenge, we devise a low-regret online recommendation algorithm, called FORM, that effectively balances the act of learning and performing fair recommendation. Our theoretical analysis confirms that FORM proficiently maintains the platform's revenue, while ensuring desired levels of fairness for both items and users. Finally, we demonstrate the efficacy of our framework and method via several case studies on real-world data.

2 - Assortment Optimization for General Multi-Purchase Choice Models

Wenhao Gu, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Anton Braverman, Tarek Abdallah

The static assortment optimization problem is a classical and well-studied problem where customers choose a single item while behaving according to some customer choice model like the ubiquitous random utility model. However, the variant where customers purchase multiple items has received less attention, primarily due to the added complexity of modeling utility-maximizing behavior over sets of items, even when considering natural extensions of the standard MNL choice model. In this paper, we propose a general multi-purchase choice model that can be viewed as a natural extension of the single-purchase utility choice models. We also study the respective assortment optimization problem without making specific distributional assumptions on the random utilities. We propose a computationally efficient algorithmic framework that is based on an asymptotic regime referred to as the large-offering regime, where the number of items available to the retailer grows large. Through this asymptotic lens, we develop an efficient approximation algorithm with corresponding asymptotic optimality guarantees under general utility distributions. Our numerical results demonstrate that our algorithm is very competitive even when the number of items available to the retailer is modest and even when the customers choose a single item.

3 - Rationality-Robust Information Design: Bayesian Persuasion Under Quantal Response

Wei Tang, Chinese University of Hong Kong, Hong Kong, Hong Kong, Yiding Feng, Chien-Ju Ho

Classic mechanism/information design imposes the assumption that agents are fully rational, meaning each of them always selects the action that maximizes her expected utility. Yet many empirical evidence suggests that human decisions may deviate from this full rationality assumption. In this work, we attempt to relax the full rationality assumption with bounded rationality. Specifically, we formulate the bounded rationality of an agent by adopting the quantal response model (McKelvey and Palfrey, 1995).

We develop a theory of rationality-robust information design in the canonical setting of Bayesian persuasion (Kamenica and Gentzkow, 2011) with binary receiver action. We first identify conditions under which the optimal signaling scheme structure for a fully rational receiver remains optimal or approximately optimal for a boundedly rational receiver. In practice, it might be costly for the designer to estimate the degree of the receiver's bounded rationality level. Motivated by this practical consideration, we then study the existence and construction of robust signaling schemes when there is uncertainty about the receiver's bounded rationality level.

4 - Too Good to Waste: Reducing Food Waste Through Markdown Platforms

Linghui Feng, Boston University, Boston, MA, United States, Pnina Feldman, Robert Swinney

This paper investigates the sustainability and efficiency of online food markdown platforms in reducing food waste. Despite established sustainability goals and efforts, significant quantities of food are still lost or wasted along the food supply chains. Several startups have developed a business model that sells surplus units from food retailers to consumers at discounted prices. Inspired by these companies, we analyze this markdown business through a stylized game-theoretical model, considering a revenue-sharing contract between a food markdown platform and a restaurant. By comparing this model with two benchmarks—no markdown and a centralized markdown approach—we find that while markdown platforms typically enhance profitability for restaurants over discarding food, they can lead to increased food waste under certain conditions. Additionally, we find that these platforms consistently perform better in reducing waste compared to the centralized markdown approach. Through a combination of analytical modeling and numerical simulations, we explore the trade-offs involved in these markdown strategies and propose an improved revenue-sharing contract to ensure sustainability. Further, in the extension of our base model, we investigate the impact of strategic consumer behaviors and market-clearing prices on restaurants' profitability and sustainability. We also discuss the potential of a government incentive program to mitigate these issues, proposing a practical design that could be implemented to realign market dynamics with environmental sustainability goals.

5 - A Modeling Framework for Tipping in the Presence of A Social Norm

Ran Snitkovsky, Tel Aviv University, Tel Aviv, Israel, Laurens Debo

Tipping is a complex phenomenon driven by social pressure, and cutting across various stakeholders -- firms, customers, and workers. Analyzing the long-run impact of policies related to tipping is therefore challenging. To facilitate such analysis, we develop a modeling framework, in which a tipping norm is formed endogenously in a market comprised of a price-setting firm offering service to potential customers. Customers choose whether to consume the service or not, and if yes, how much to tip the server afterwards. With tipping, customers show reciprocity to the server by sharing a fraction of their surplus, but also undergo social pressure to comply with the prevailing norm of tipping. This tipping norm is shown to evolve endogenously through a dynamic process of sequential market adjustments over time: the average tip
determines in each time period the tipping norm for the next period, causing the firm to adapt its price and customers to adapt their tips

accordingly. The average tip-to-price ratio converges (from below) to a stable market-equilibrium outcome. Characterization of this equilibrium outcome allows us to derive qualitative results about the long-run impact on tipping of different exogenous factors such as customers' sensitivity to social pressure and the quality of service. Building on this framework, we further investigate several economic implications of tipping pertaining social welfare, labor cost, and service quality, thus, uncovering incentives and trade-offs to which the tipping mechanism give rise, from the firm’s, the worker’s, and the customer’s perspectives.

TC17

Summit - 337

Socially Responsible Supply Chains

Invited Session

MSOM: Supply Chain

Chair: Gizem Korpeoglu, TU/e, Eindhoven, 5632TE, Netherlands

Co-Chair: Sina Khorasani, Vanderbilt University, Franklin, TN, United States

1 - Mergers of Consumer Cooperatives

Cigdem Gizem Korpeoglu, TU/e, Eindhoven, Netherlands, Sina Khorasani

Consumer cooperatives are consumer-owned and managed enterprises that aim to achieve buyer power and maximize their members' welfare. Recently, several cooperatives in major economies, such as the UK and Italy, merged to increase their buyer power and achieve lower prices for their members. We seek to understand how these mergers affect strategic interactions among supply chain players, market outcomes, and consumer welfare. We consider two-tier supply chains where multiple consumer cooperatives procure from a market for their member consumers. Multiple suppliers produce for this market and can increase their supplies by incurring scale-up costs. We show that mergers of cooperatives reduce prices, as intended. This enables consumers to allocate more of their income to purchasing other goods. However, lower prices also induce suppliers to reduce their production quantities, thereby causing each consumer to receive less of the supplied product. We find that this underproduction is even more pronounced in industries with low production scale-up costs. Due to this tradeoff between lower prices and underproduction, mergers of cooperatives make a nuanced impact on consumer welfare. We show, interestingly, that mergers harm both member and non-member consumers when the pre-merger number of cooperatives or scale-up cost is below a certain threshold. Otherwise, mergers benefit all consumers. We expand our results by considering horizontal and vertical differentiation among cooperatives and show that our results are robust and that differentiation increases the benefit of mergers. Policymakers should maintain healthy competition among cooperatives to maximize consumer welfare, especially in markets with low production scale-up costs.

2 - Supplier-Initiated and Buyer-Imposed Social Responsibility Codes of Conduct

Sriram Narayanan, Michigan State University, East Lansing, MI, United States, Han Zhang, Fei Gao, Mevan Jayasinghe

Buyer-imposed social responsibility codes of conduct have led to limited improvement in the social responsibility practices of suppliers. In response supplier-initiated voluntary codes of conduct have been increasingly accepted. We study a supplier’s decision on whether to adopt a voluntary code and a buyer’s decision on whether to impose its own code. We find that the stringency of the voluntary code determines whether it can effectively signal social responsibility. When the voluntary code signals ineffectively the buyer may impose its own code. Indeed the buyer may impose its code on a supplier that has adopted the voluntary code yet refrain from imposing its code on a supplier without any code. The strategy creates two distinct classes of suppliers: premium and economical. In general the voluntary code is not a substitute for the buyer’s code: the buyer may scrutinize a supplier more closely when the supplier has adopted the voluntary code. When the buyer incurs a low cost for the publicity of a supplier violation, increasing the stringency of the voluntary code will dampen supplier adoption of the code and loosen monitoring by the buyer at the same time, potentially increasing the likelihood of a violation slipping through detection.

Keywords: supplier code of conduct, social responsibility, signaling game, asymmetric information

3 - The Value of Negative Self-Disclosure in Supply Chain Social Responsibility

Tao Lu, University of Connecticut, Storrs, CT, United States, Yang Tong, Shi Chen

Some firms self-disclose social responsibility (SR) violations discovered in their supply chains. Motivated by this fact, we develop a game-theoretical model to explore the potential benefit of self-reporting negative SR incidences. We consider a buyer sourcing from an existing supplier and an incoming supplier. The buyer can be of high or low SR type, which is asymmetric information to the incoming supplier. First, the buyer audits the existing supplier, and if a violation is discovered, decides whether to disclose the violation. Then, the buyer audits the incoming supplier, whereas the incoming supplier chooses an SR effort to qualify for the buyer’s supply base. We show that through negative disclosure, a high-type buyer can effectively signal its ethical level to the incoming supplier, and such a separating equilibrium occurs if and only if the audit cost is within a medium range. Our results provide insights into how the audit cost and accuracy and the cost of negative disclosure influence an ethical firm's self-disclosure strategy and the SR violation risk in a supply chain.

4 - How Do Producers Fare with Fair Trade?

Ying Zhang, Santa Clara University, Santa Clara, CA, United States, Yen-Ting (Daniel) Lin, Adem Orsdemir, Heng Chen

Fair trade certificate promotes a sustainable livelihood for producers in developing countries. It ensures that producers are paid at a fair wage and potentially protects producers from the volatility in the commodity market. In this paper, we examine the impact of fair trade certificate on a fair trade retailer’s decisions, profitability and participating producers’ welfare. We also explore the impact of market size and proportion of myopic farmers on the optimal fair trade price that maximizes the producers' welfare. Furthermore, we investigate the disparity issue between myopic farmers and strategic farmers.

TC18

Summit - 338

Emerging Technologies in Online Platforms

Invited Session

Revenue Management and Pricing

Chair: Bing Bai, McGill University, Montreal, QC, Canada

1 - Paying with TIME OR with Money: What Drives Customer Waiting Preferences?

Evgeny Kagan, Johns Hopkins Carey Business School, Baltimore, MD, United States, Andrew Davis, Kyle Hyndman

Many service offerings are characterized by two attributes: the price of the service and the time it takes to access or receive it. We experimentally study people’s preferences among time-money bundles. Participants in our experiments made repeated binary choices between two waiting formats: a short wait with a small monetary payoff, and a longer, potentially stochastic wait with a larger monetary payoff. We found that (1) certain waits are preferred to uncertain waits, though small amounts of uncertainty are ignored (2) there are no significant differences in how people evaluate ambiguous waits (where the probabilities of duration outcomes are unknown) compared to uncertain waits (where probabilities are known), (3) two-point waits (Bernoulli distribution) are preferred to continuous waits, while for continuous distributions, the upper bound may matter more than the variance and (4) people are most interested in receiving right-tail information about the duration of a wait, and most strongly respond to changes in the right tail.

2 - Digital Recommendations to Guide Cultural Experiences: Experimental Evidence at the Van Gogh Museum

Abhishek Deshmane, Georgia Institute of Technology, Atlanta, GA, United States, Ali Aouad, Victor Martínez de Albéniz

Technology interfaces and digital platforms play an increasingly vital role in mediating services and customers, offering relevant recommendations, context, and amplifying experiences. Through a field experiment conducted across the years 2022-2024, this study evaluates the impact of multimedia guide offerings at the Van Gogh Museum, focusing on visitor engagement, congestion, and educational objectives of the museum. Our findings aim to evaluate the effectiveness of such technology systems in cultural contexts by guiding visitors to specific exhibits, segmenting visitors based on their diverse needs and preferences, and addressing the challenges in estimating the marginal value of expanding tour offerings. Additionally, we explore the impact of visitor fatigue on the quality of the experience and propose leveraging experimental methods to optimize tour content for maximum audience engagement.

3 - The Value of Logistic Flexibility in E-Commerce

Bing Bai, McGill University, Montreal, QC, Canada, Tat Chan, Dennis Zhang, Fuqiang Zhang, Yujie Chen, Haoyuan Hu

In recent years, improving shipping speed in online retailing has faced significant diminishing marginal returns and significant cost increases. As a consequence, many online retailers have started to explore improving other aspects of the shipping experience to attract customers. In this paper, we use the introduction of local pick-up stations by Alibaba to study the impact of improving logistic flexibility on online retailing. We develop a choice model to study how logistic flexibility affects customers' purchase decisions. We find that customers value two types of logistic flexibility -- the value of time flexibility, which offers them the flexibility to receive their packages in their most preferred time; and the value of choice flexibility, which offers them the flexibility to make pick-up choices until the last minute to accommodate time uncertainty.

4 - Omnichannel Operations in on-Demand Delivery Platform with Buy-Online-and-Pick-up-in-Store

YU GUO, University of Science and Technology of China, Hefei, China, People's Republic of, Fei Gao, Wenchang Zhang, Liu Ming

On-demand delivery platforms now provide consumers with the option to buy online and pick up in-store (BOPS). We examine such an omnichannel on-demand economy, where consumers decide how their orders are delivered: by independent couriers or by themselves. First, we develop a stylized model to study the impact of BOPS on platform pricing strategies on both the courier and consumer sides. Our findings reveal new insights that differ from the delivery-only setting. After implementing BOPS, (i) the demand for delivery service decreases, yet the platform may increase price in the delivery channel; (ii) the platform ``recruits'' consumers to fulfill their own orders, but it may charge BOPS consumers an even higher price than for deliveries; (iii) surge pricing may no longer be optimal in response to a demand surge. Furthermore, the model is extended to incorporate a government-imposed wage floor. Although implementing BOPS reduces courier welfare, we show that policymakers may need to lower the wage floor to increase delivery opportunities for couriers. Finally, we estimate the model primitives using data from a leading meal-delivery platform in China and employ the estimated model to predict the impact of BOPS and wage regulations. We find that BOPS enhances platform profit by 12.00%, whereas it reduces labor welfare by 50.95%. Nevertheless, introducing wage floors alongside BOPS can improve labor welfare more effectively (up to +278.44%) compared to without BOPS (up to +220.14%). Moreover, with BOPS, a small wage floor can significantly increase labor welfare while causing minimal harm to the platform's profit.

5 - Optimal Subsidy Design: Application to Staple Food Subsidy

Zhicong Hu, London Business School, London, United Kingdom, Ali Aouad, Kamalini Ramdas, Alp Sungu

Designing subsidy programs presents a significant challenge: balancing policy efficiency with the flexibility of individual choice. “Paternalistic” approaches prioritize social outcomes, such as promoting the consumption of nutritious but less popular foods. This focus, however, can come at the expense of individual preferences, potentially leading to low uptake of the program by the intended beneficiaries. Conversely, policies that heavily weight individual choice may be inefficient. Empirical findings from a randomized controlled trial (RCT) on staple food subsidies in a low-income community illustrate this trade-off. In response, we propose an optimization model for subsidy design based on multinomial logit preferences, incorporating both in-kind subsidies and cash transfers. By leveraging technical insights from the revenue management literature, we identify the structure of optimal policies and develop practical solution methods. We calibrate our model using RCT data to generate counterfactual policies along the Pareto Frontier, aiming to balance efficient social outcomes with reduced paternalism.

TC19

Summit - 339

New Methods for Inventory Policies

Contributed Session

Chair: Reuben Joseph, Binghamton University (SUNY)

1 - Advertising Selling Service and Selling Format in Online Retailing

JIANYUE WANG, The Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Ki Ling Cheung, Albert Ha

We develop several game theoretic models to study the selling format and seller service strategies of an online retailer. We consider two common selling formats, agency selling and reselling, and one common seller services, advertising selling. When the online retailer offers advertising selling service, there is a cost sharing flexibility effect that creates more incentive for a seller to buy advertising under agency selling than under reselling. Moreover, a higher cost of advertising effort mitigates the double marginalization effect of wholesale price under reselling. We fully characterize the firms' equilibrium decisions. The online retailer prefers agency selling to reselling if the commission rate, demand signal accuracy or demand uncertainty is high, or the cost of advertising effort is low. Moreover, the online retailer prefers to sell advertising servcie if the commission rate, demand signal accuracy, demand uncertainty or cost of advertising effort is high. We show the impact of offering advertising selling on the profits of the seller and the supply chain. When more seller services are offered, the parameter space in which the equilibrium selling format is agency selling (respectively reselling) becomes larger (respectively smaller). Moreover, offering advertising service has a wider positive impact and a narrower negative impact when another seller service is available than when it is not available. The model is extended by considering the seller determines the selling format or including another seller service, information sharing service.

2 - Game-Theoretic Integration of Red Team Survey Data in Multi-Layer Security Systems

Ian Unson, University at Buffalo, Buffalo, NY, United States, Jun Zhuang

With the rise of domestic terrorism and soft target attacks, it is crucial we optimally allocate defensive resources across multiple security layers in venues such as airports, subway stations, sports venues and houses of worship. By allocating resources across these security layers properly we will aim to minimize the damages caused by a potential attacker or even be able to deter them. The attacker on the other hand, has the ability to observe how the defender allocates their resources and tries to circumvent the defenses we have laid out in order to maximize their objective. Building upon previous research in game theory, resource allocation problems and sociological surveys of participants playing the role of strategic attackers, we look at methods of updating the defenders' best response given updated attacker strategies. These attacker models have been built upon red teaming survey data collected from studies where participants played the role of an attacker. These attacker models not only take into account features within the game such as cost, payoffs, probability of success, etc. but also take into account characteristics of the attackers, such as numeracy skills and risk preferences. By utilizing these updated attacker models to solve for better Nash equilibria we aim to provide defenders with better decision-making tools that can deter or reduce the damages caused by soft target attacks.

3 - Dynamic Estimation of Random Effective Shelf Lives and Policy Optimization for Perishables

Ulku Gurler, Bilkent University, Ankara, Turkey

In multi-echelon supply chains, it is common that random demand and batch ordering processes result in effectively random shelf lives even when items have fixed lifetimes as issued by the uppermost echelon. Storage conditions at each location may further exacerbate this stochasticity rendering true remaining shelf lives unobservable. In this talk, I consider the dynamic estimation of effective shelf lives and optimization of simple, reasonable inventory control policies. Focusing on a single-location perishable inventory system in isolation as decoupled from the rest of the supply chain, an estimation procedure in the presence of censored shelf life observations and a heuristic policy optimization is provided. Effects of system parameters on estimation performance and relevant cost measures are discussed through numerical examples.

4 - Decision Support System with Integrated Ensemble Forecasting for Dynamic Inventory Control using System Dynamics

Reuben Joseph, Binghamton University (SUNY), Binghamton, NY, United States, Kunal Mangaonkar, Mohammad Khasawneh, Sreenath Chalil Madathil

Hospital inventory management ensures the continuous availability of medical supplies while aiming to reduce costs. Traditional methods, relying on heuristic rules and personal experience, often lead to overstocking, understocking, variable lead times, and inconsistent ordering processes that do not align with Group Purchasing Organization (GPO) and local contracts, resulting in higher prices.

This paper presents the Dynamic Inventory Optimization (DIO) model, a comprehensive solution to the challenges of hospital inventory management. The model tackles irregular inventory levels, inefficient ordering processes, high stock-outs, and low fill-rates. It incorporates a multi-type and multi-step time series forecasting and system dynamics framework known as the Dynamic Drum-Buffer-Rope (DDBR) replenishment approach. The model's effectiveness is further enhanced by using stacking ensemble learning models, including XGBRegressor, LSTM, Random Forest Regressor, and a Meta layer. These models outperform traditional algorithms in predicting future trends. The DIO model also introduces the innovative use of the Powell search Gradient descent algorithm, which dynamically adjusts inventory levels based on real-time data and usage trends. This approach prevents stock-outs, reduces inventory excess, and optimizes replenishment to reflect actual consumption patterns, significantly reducing costs and improving supply chain performance. Rigorous experimentation and simulation have validated the model's effectiveness, presenting a transformative tool for strategic stock allocation and decision support in hospital settings. Future work will focus on integrating the DIO model into real hospital ERP systems to generate real-time automated replenishment orders based on consumption trends with minimal manual interaction.

Keywords: hospital inventory optimization; system dynamics; Powell search algorithm; ensemble time-series forecasting

TC20

Summit - 340

Future Directions in ML and Optimization for Resilience

Panel Session

Decision Analysis Society

Co-Chair: Himadri Sen Gupta, University of Oklahoma, 202 W Boyd St, Norman, OK, 73019, United States

1 - Moderator Panelist

Himadri Sen Gupta, University of Oklahoma, Norman, OK, United States

2 - Panelist

Christopher Zobel, Virginia Tech, Blacksburg, VA, United States

3 - Panelist

Jose Ramirez-Marquez, Stevens Institute of Technology, Hoboken, NJ, United States

4 - Panelist

Gabriela Gongora-Svartzman, Carnegie Mellon University, Pittsburgh, PA, United States

5 - Panelist

Charles Nicholson, The University of Oklahoma, Norman, OK, United States

TC21

Summit - 341

Decision Science in Transportation Research

Invited Session

Decision Analysis Society

Chair: Mingwei Guo, North Dakota State University, Fargo, ND, United States

1 - Probabilistic Modeling for Uncertainty Quantification in Oil Pipeline Predictions

Bright Awuku, North Dakota State University, Fargo, ND, United States, Ying Huang, Nita Yodo, Eric Asa

Pipelines are critical infrastructure for transporting hydrocarbons, requiring massive capital investments. However, they are susceptible to failures that can lead to economic losses and environmental damage. Pipeline management agencies rely on predictions of failure events to plan maintenance, rehabilitation, and repair (MR&R) activities. State-of-the-art techniques yield point estimates without quantifying the uncertainty in failure-cause predictions, limiting their reliability for risk-based maintenance decisions. Reliable uncertainty quantification (UQ) of model predictions is essential for effective data-driven decision-making. While machine learning (ML) models do not inherently provide well-calibrated uncertainty estimates, more rigorous Bayesian techniques exist but are computationally expensive coupled with restrictive data assumptions. This paper proposes a conformal prediction methodology for assessing prediction uncertainty in ML-based oil pipeline failure causes, addressing the need for accessible and practical UQ techniques. The approach provides prediction regions with valid coverage probabilities, enabling quantification of uncertainties associated with failure-cause predictions to support risk-informed oil pipeline maintenance planning. The proposed methodology has implications for transportation infrastructure management by providing a framework to quantify uncertainties in condition predictions for assets like bridges, pavements, and railroads. The proposed method enhances confidence in using data-driven models for making informed MR&R decisions in critical transportation applications. This work contributes to an uncertainty estimation approach that overcomes the limitations of current techniques while being efficiently applicable to various ML model architectures without restrictive assumptions. The ability to generate valid prediction intervals enriches predictive analytics for more informed decision-making in asset management.

Acknowledgment

Authors grateful for NSF EPSCoR grant OIA-2119691 funding this study.

2 - Investigation of Condition Assessment Aiding Pavement Rehabilitation Decision-Making

Kathryn Quenette, North Dakota State University, Fargo, ND, United States, Xinyi Yang, Luyang Xu, Xingyu Wang, Ying Huang

The number of registered vehicles in the U.S. continues to rise, reaching approximately 282 million in 2021. This increase heightens concerns about detecting and repairing cracks before they become severe, since repairs can significantly extend the service life of the pavement at a low cost. To achieve this maintenance strategy, one solution is real-time monitoring, which tracks crack development to inform repairs. However, this method demands substantial manpower and traditional techniques typically only detect surface cracks, not those developing from the bottom up. To address these challenges, this study introduces a sensor system installed beneath the pavement, specifically using Fiber Bragg Grating (FBG) and distributed fiber optical sensors, embedded within the asphalt layers and between the asphalt and concrete

layers. This study aims to embed sensors beneath highway pavement to collect data for traffic analysis and pavement condition assessment, ultimately developing a new method for pavement integrity management planning. This proposed system would inform engineers when pavement replacement or crack sealing is necessary and allow users to monitor traffic and identify high-risk roads. Integrating sensors and creating this database would enable more efficient data access and informed pavement rehabilitation decisions.

3 - Empowering Super-stars or Rising-stars? Dynamics of the High-Speed Rail Network in China

Yucong Li, Tongji University, Shanghai, China, People's Republic of, Dongping Cao, Guangbin wang

The development of high-speed rail (HSR) systems as a critical component of the global 'infrastructure turn' is deeply embedded in complex regional socioeconomic contexts. Drawing on advances in actor-oriented network modelling and longitudinal inter-city HSR data in China during 2008-2022, this study explores how the evolution of the HSR network is dynamically and differentially underpinned by underlying attribute-based effects of regional socioeconomic scale and growth rate to balance the short-term efficiency and long-term equity of nationwide development. After controlling for related structure-based effects (e.g., transitive triads and balance) as well as attribute-based effects related to geographic, natural and cultural contexts, the results provide evidence that while the scale and the growth rate of regional economy and population both positively impact HSR tie formation, the impacts of attribute-based effects related to scale significantly diminish over time whereas the impacts of effects related to growth rate exhibit a significant strengthening trend. The results further illustrate that while the scale-related and growth-rate-related effects of regional economy function as distinct substitutes in driving the formation of HSR ties, the two corresponding effects of regional population significantly complement each other, providing clear evidence for the heterogeneous roles of regional economic and social contexts in underpinning HSR network dynamics. The findings provide implications for more effective and equitable regional socioeconomic development through strategic inter-regional HSR network planning.

4 - Rookie Routing and Other Simulated Submodel for Navigation Systems

Mingwei Guo, North Dakota State University, Fargo, ND, United States

This research aims to enhance road safety by providing alternative guidance tailored specifically for new and inexperienced drivers. By analyzing their driving behaviors, the study seeks to integrate adaptive navigation systems that offer real-time, context-sensitive assistance. The goal is to develop a guidance system that reduces stress and improves decision-making for rookie drivers, ultimately leading to safer and more confident driving experiences.

TC22

Summit - 342

Data-Enabled Decision-Making in Energy Systems and Markets

Invited Session

Decision Analysis Society

Chair: Zhirui Liang, Johns Hopkins University, Baltimore, MD, United States

Co-Chair: Yury Dvorkin, Johns Hopkins University, Baltimore, MD, United States

1 - Dynamically Learning Incentives for Load Control

Joshua Comden, NREL, Golden, CO, United States, Andrey Bernstein, Guido Cavraro

With the increase in variable renewable resources, and the electrification of many industries, it is becoming increasingly important to take advantage of flexible loads as controllable resources. However, incentivizing owners of flexible loads to relinquish some control in an effective and cost-efficient way has remained elusive. In this talk, we propose the use of a model-free feedback-based control algorithm that adjusts incentives for each flexible load owner based on the benefit to the system. One major advantage to adjusting incentives based on system benefit is that it rewards overall behavior instead of adjustments from a baseline which are difficult to accurately estimate in practice.

2 - A Contextual Bandit Approach for Value-Oriented Prediction Interval Forecasting

Yufan Zhang, University of California San Diego, San Diego, CA, United States, Honglin Wen, Qiuwei Wu

Prediction interval (PI) is an effective tool to quantify uncertainty and usually serves as an input to downstream robust optimization. Traditional approaches focus on improving the quality of PI in the view of statistical scores and assume the quality improvement will lead to a higher value in the power systems operation. However, such an assumption cannot always hold in practice. In this paper, we propose a value-oriented PI forecasting approach, which aims at reducing operational costs in downstream operations. For that, it is required to issue PIs with the guidance of operational costs in robust optimization, which is addressed within the contextual bandit framework here. Concretely, the agent is used to select the optimal quantile proportion, while the environment reveals the costs in operations as rewards to the agent. As such, the agent can learn the policy of quantile proportion selection for minimizing the operational cost. The numerical study regarding the day-ahead and real-time operation of a virtual power plant verifies the superiority of the proposed approach in terms of operational value. It is especially evident in the context of extensive penetration of wind power.

3 - From Sensing to Control: Strategies for Real-TIME Model Identification in Smart Grids

Vivek Khatana, University of Minnesota, Twin Cities (Minneapolis, MN, US), Minneapolis, MN, United States, Chin-Yao Chang

In this talk, I will focus on the problem of system model identification in power distribution networks through an innovative agent-based approach. The proposed method emphasizes distributed and localized model configurations to enhance robustness and address prevalent mislabeling issues within distribution grids. Departing from conventional practices of relying on historical data for offline model identification, the strategy presented in the talk introduce online model updates utilizing real-time data by employing only the latest measurements. This methodology offers significant advantages, including a large reduction in the communication network’s bandwidth requirements by minimizing the data exchanged at each iteration and enabling the model to adapt in real-time to disturbances. Furthermore, we extend the state-of-the-art from linear frameworks to more complex non-linear convex models. This extension is validated through numerical studies demonstrating improved voltage regulation performance. These advancements aim to fully exploit the capabilities of newly installed sensors in power distribution grids, significantly enhancing their practical value and utility.

4 - Dynamic Net Metering for Energy Communities

Ahmed Alahmed, Cornell University, Ithaca, NY, United States

We propose a grid-aware and social welfare maximizing market mechanism for an energy community that aggregates individual and community-shared energy resources under a general net energy metering (NEM) policy. Referred to as Dynamic NEM (D-NEM), the proposed mechanism dynamically charges/rewards community prosumers based on a two-part pricing rule that has a threshold-based structure. By formulating the decisions of the community manager and its prosumers as a Stackelberg game, we show that the proposed two-part pricing achieves a Nash equilibrium and maximizes the community's social welfare in a decentralized fashion while ensuring that prosumers' optimal decisions abide by the grid constraints. We prove that D-NEM is a stabilizing mechanism, as it achieves both individual and group rationalities. Finally, we argue that D-NEM is a just and fair mechanism to all community prosumers by showing that it conforms with the axioms of the cost-causation principle.

5 - Time series aggregation in energy system models with temporal and network constraints

Sonja Wogrin, Graz University of Technology, Graz, Austria, David Cardona-Vasquez

In this talk we discuss how to improve computational tractability of large-scale Linear Programs (applied to energy system model) that have network constraints and temporally-linked constraints, such as ramping constraints. We do this via a method - developed by us - that is referred to as basis-oriented aggregation. We present an extension of this method using machine learning, which allows for: decoupling of the large LPs, parallelization and in turn faster resolution times.

TC23

Summit - 343

Optimizing Industrial and Construction Supply Chains

Contributed Session

Chair: Songi Kim, Inha University, 100 Inha-ro, Michuhol-gu, 22212

1 - Enhancing Safety Management in Circular Supply Chains through Blockchain Integration: A Stakeholder Perspective

Zhuowen Chen, Worcester Polytechnic Institute, Worcester, MA, United States, Joseph Sarkis, Yanji Duan, Abdullah YILDIZBASI

Safety management in circular supply chain (CSC) processes is critical yet insufficiently explored in operations management. The complexity of circular practices such as reuse, repurpose, remanufacture, and recycling raises challenges in maintaining safety standards. These challenges call for innovative management solutions to ensure effective safety protocols. Blockchain technology, recognized for its transparency and traceability, emerges as a promising tool for enhancing safety management across CSCs. The decentralized and immutable characteristics of blockchain make it particularly suitable for ensuring the integrity and reliability of safety data across the CSC. Stakeholder attitudes and perceptions play a pivotal role in the initial adoption of blockchain for CSC safety management. This study utilized a discrete choice experiment to explore the perceptions and willingness-to-pay among various stakeholders involved in CSCs. The methodology focuses on understanding stakeholder preferences concerning blockchain integration and their readiness to adopt blockchain-based safety management systems. The attributes examined focus on blockchain capabilities, safety and circularity practices within the supply chain. The findings of this study provide significant insights into the feasibility and desirability of blockchain applications for enhancing safety within CSCs. Moreover, the paper discusses the implications for future research and practice management of safety in CSCs. It suggests areas where further investigation is needed and highlights how managers can implement blockchain solutions in circular supply chains. These insights aim to contribute to the broader discourse on pursuing circular supply chains through innovation technologies.

2 - Operational Heterogeneity in U.S. Blast Furnaces

Elina Hoffmann, Carnegie Mellon University, Pittsburgh, PA, United States, Valerie Karplus, Chris Pistorius

Steel is vital to modern economies; however, its production is highly energy- and emissions- intensive. It is important to decouple steel production from CO_2 emissions for the sector to achieve its climate goals. While recent studies have focused on low-carbon steelmaking technologies, little attention has been paid to the CO_2 emissions reduction available with existing production processes. Operating existing technologies in a manner that minimizes CO_2 emissions can offer meaningful reductions in the near- and medium- term without the large capital outlays required of net-zero compatible routes. This study investigates the operational heterogeneity within- and across- a single steelmaking process technology, focusing on the blast furnace. Leveraging annual data on U.S. blast furnaces, we find considerable variation in operation resulting in direct emissions intensity estimates ranging from 1.32 to 1.67 tons of CO_2 per ton hot metal (tCO_2/tHM). The U.S. fleet may have been able to reduce cumulative CO_2 emissions more than 13% in the past eleven years had furnaces been operating at their most CO_2 efficient. We leverage additional datasets with high facility and temporal resolution to investigate how data aggregation may influence CO_2 emissions patterns, and in turn the impact of different decarbonization policies. Our work suggests that a global, 'best available technology', or even country-specific emissions factor fails to capture the heterogeneity in blast furnace operation. Policies incentivizing technologies to operate in a way that minimizes their CO_2 emissions may result in considerable emissions reductions that may not be realized when treating technology as a static monolith.

3 - Scheduling The Production of Prefabrication Construction Supply Chains Considering Variable Delivery Due Dates

Yong-Woo Kim, University of Washington, Seattle, WA, United States, Chung Ho, Zelda Zabinsky

In recent years, the construction industry has seen a significant increase in the application of prefabrication and modularization. Improving the efficiency on the operation and reliability on the delivery of prefabrication supply chains becomes critical for construction projects. However, the fluctuating nature of construction schedules poses challenges for the planning of prefabrication supply chains and adaptations when changes occur. The experience during the COVID-19 pandemic highlighted the impact of unavailable and delayed materials on construction projects, and the cancellation of construction projects on the supply chain. Our study focuses on prefabrication supply chains with multiple fabrication shops that consider the uncertainty of required delivery due dates to construction sites. An industry survey carried out by the authors reveals that changing the required delivery due dates of prefabricated products is common in construction projects. This presentation introduces a multi-objective stochastic programing model that facilitates the production planning of prefabrication products at multiple fabrication shops considering variable delivery times. The scenario-based solution balances minimal total cost with minimal total delay. Additionally, computational results are presented for an example problem with three fabrication shops, ten construction projects and fifteen products. The model allows fabricators to develop a robust set of optimal schedules that are capable of flexibly balancing cost and time reduction objectives.

4 - Optimizing a rolling process for manufacturing small steel bars using CART

Hyundeok Jang, Sungkyunkwan University, Suwon-si, Korea, Republic of, KwangHo Jeong, Dong-Hee Lee, Sung-Jun Lim, Sang-Hyeon Lee, Jong-Eun Lee

Small steel bars (SSBs) are common steel products manufactured through a rolling process. Surface defects on SSBs, which are critical quality problems of SSBs, are often incurred during the rolling process. In order to prevent the surface defects, key process variables affecting the surface defects should be identified and controlled. Recently, it has been common to collect a large amount of operational data from the rolling process which might include valuable information for improving the rolling process. This study attempts to obtain an optimal condition of rolling process variables where the surface defects are minimized by using a particular data mining strategy called classification and regression tree (CART). By analyzing node information from CART, which includes both leaf nodes and their ancestors, it is possible to identify several candidates for the optimal condition. These candidates are represented in hyperrectangle forms, demonstrating this approach's robustness and potential for substantial improvements in minimizing the surface defects.

5 - Service provider recommendation model for shared additive manufacturing supply chain: a case study

Songi Kim, Inha University, Incheon, Korea, Republic of, Kyung-A Kim, Hosang Jung

In the shared manufacturing platform, self-organized individuals can participate in manufacturing activities as service providers. When integrated with additive manufacturing (AM) technologies, shared manufacturing holds significant potential to produce large amounts of one-of-a-kind products. However, despite its potential, the concept still remains underexplored due to a scarcity of real-world examples. This study begins by analyzing the open-source assistive technology-sharing platform in South Korea, focusing on two main participant groups: people who need assistive technologies (such as older adults and people with disabilities) and people who contribute new designs of assistive technologies. It identifies the unmet needs of the two participant groups within this open-source platform. Building upon the analysis, we propose a shared AM supply chain that introduces a third participant type: manufacturers with AM machines who are willing to share their manufacturing capabilities and resources to produce assistive devices based on open-source designs. Subsequently, we develop the manufacturer recommendation model that provides the best AM manufacturer candidate to the customer, considering the manufacturing requirements of the assistive device, the customer's preferences between lead time and price, and the capabilities of the manufacturers. Finally, we apply the recommendation model to the case study with the example assistive device and potential manufacturers in Incheon, South Korea.

TC24

Summit - 344

Bike Sharing and Urban Transportation

Contributed Session

Chair: Dimitrios Letsios, King's College London, London, United Kingdom

1 - Park Smart or Face the Music: Understanding Users' Orderly Parking Behavior of Dockless Shared Bikes from the Perspective of Deterrence Theory

Jianliang Hao, California State University, Chico, Chico, CA, United States, Zhenya Tang

The rapid growth of dockless bike-sharing has revolutionized urban transportation. However, the disorderly parking of dockless shared bikes has emerged as a critical challenge, significantly hindering the benefits of this system by causing obstructions, safety hazards, and public nuisance. This study aims to investigate users' orderly parking behavior of dockless shared bikes through the lens of deterrence theory. Our findings reveal that punishment severity, punishment certainty, personal norm, and descriptive norm have significant positive effects on users' attitudes towards orderly parking, which in turn positively influences their orderly parking behavior. This study contributes to the literature by extending the application of deterrence theory to the context of dockless bike-sharing and providing empirical evidence on the role of deterrence factors in shaping users' orderly parking behavior. Our findings offer valuable insights for bike-sharing operators and policymakers to develop effective strategies.

2 - The Paradox of Choice: Why Less is More for Crowdsourcing in Bike Sharing System

jinjia huang, Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Zhenzhen Yan, Mabel C. Chou, Chung Piaw Teo

Inspired by New York's Citi Bike Bike Angels Program and Boston's Blue Bikes, we investigate the use of registered volunteers to reposition empty bikes for riders in a bike-sharing system. In Bike Angel programs or similar initiatives, volunteers move bikes from stations with excess bikes to stations in need of bikes in exchange for rewards. To reduce redundant re-balancing moves by volunteers, we propose a method that restricts the set of routes eligible for rewards. This demonstrates that limiting the deployment options for volunteers can actually make re-balancing activities more effective. Leveraging the modeling power of copositive cone programming, we develop a novel network design model tailored for re-balancing operations in crowdsourcing environments. Our model explicitly integrates critical factors such as volunteer availability, choice behaviors, and supply-demand variations, setting our work apart from traditional approaches in process flexibility and most existing studies. Our simulation experiments show that our approach provides a highly effective and cost-efficient solution to mitigate re-balancing issues in these systems.

3 - Vehicle Electrification and Grid Impact Assessment Tool for Informed Charging Infrastructure Expansion

Farhad Angizeh, Massachusetts Institute of Technology, Cambridge, MA, United States, Jessika E. Trancik

This paper reports on our three-layer electric vehicle (EV) charging infrastructure and grid impact assessment tool to aid policy stakeholders and developers in evaluating expansion strategies for widespread EV adoption. The first layer employs our *TripEnergy* model to estimate the energy needs for diverse personal vehicles, considering longitudinal travel patterns and driving behaviors. The second layer quantifies time-resolved energy needs to explore potential EV charging solutions (e.g., at home, workplace, public, and highway), guiding strategic charging infrastructure expansion. This layer also examines the role of behavioral modification in further enabling EV adoption. The third layer investigates the spatiotemporal grid impacts, informs project siting, and addresses whether upgrades are needed for distribution feeders and substation equipment to meet peak demands. Several case studies are demonstrated to highlight the key determinants of effective policy-making strategies and infrastructure expansion investments to support vehicle electrification.

4 - Static and Dynamic Inventory Rebalancing for Optimizing Service Loss in Bike Sharing

Dimitrios Letsios, King's College London, London, United Kingdom, Haoxiang Wang, Xinyi Ye

We elaborate on inventory rebalancing in bike sharing. The goal is to decide how to reposition bikes between the stations of a bike sharing network so as to optimize service loss. Previous work computationally demonstrates the benefits of dynamic compared to static rebalancing. We propose new optimal algorithms for both problem variants. On the negative side, we show that static rebalancing results in O(1/epsilon) proportion of successful requests. On the positive side, we prove that dynamic rebalancing achieves an Omega(1-epsilon) performance guarantee. To our knowledge, this is the first theoretical characterization of the dynamic rebalancing benefits compared to static rebalancing. Next, we consider the more general problem of optimizing service loss with bounded rebalancing costs. For both the static and dynamic variants, we develop new scalable mixed-integer approaches. We computationally substantiate the proposed approaches using data from the London bike sharing network.

5 - Repositioning in bike sharing systems with broken bikes considering on-site repairs

RUNQIU HU, The University of Hong Kong, Hong Kong, Hong Kong, WY Szeto

On-site repairs by repairmen can handle the broken bikes in a bike-sharing system and satisfy the usable bike demand simultaneously without vehicle repositioning for broken bike collection for off-site repairs. However, such option has not been jointly considered with vehicle-based repositioning for both usable and broken bikes in the literature.

This paper presents a mixed integer linear programming model for a static bike repositioning problem with broken bikes combining vehicle-based bike delivery/collection and labor-based on-site repairs. The model aims to minimize the total cost of user dissatisfaction and the carbon emissions generated during the repositioning. Two model variants are also developed for scenario comparison. An algorithm combining the hybrid genetic algorithm with a greedy heuristic was proposed to solve the problem. Results show introducing an on-site repairman can lead to lower user dissatisfaction but more carbon emissions in some cases. Moreover, the introduction of a repairman proves effective in reducing the total cost of user dissatisfaction and carbon emissions when the broken bike inventories at deficit stations are approximately half of the usable bike shortages. With an on-site repairman, the lowest total cost and carbon emissions occur when the broken bike quantity is exactly the usable bike shortage. After the number of broken bikes exceeds that shortage, rebounds in the total cost and carbon emissions are observed. Furthermore, separating the routes of the repairman from the truck can reduce the total cost of user dissatisfaction and carbon emissions, with decreased user dissatisfaction but occasionally slightly higher carbon emissions.

TC25

Summit - 345

FICO/Amazon

Invited Session

Technology Showcase

1 - From Data to Decisions with the Help of Gen AI and Great Optimization Technology

Dinakar Gade, FICO, Bozeman, MT, United States, Oliver Bastert

You have a team with a great analytics background. They have created advanced analytical tools using Python, R, or your current optimization solver. They managed to derive insights from your data and figured out what the best path forward is. But the tools remain cumbersome to use and are not accessible by business domain experts and business users. It is time to put these critical analytical insights into the hands of the business – with the help LLMs and great optimization technology.

In this tutorial, you'll learn how FICO's Xpress Optimization capabilities make it possible to embed your analytic and optimization models within business user-friendly applications. See how to supercharge your models with simulation, fast and robust optimization, reporting, what-if analysis, and agile extensibility for your ever-changing business. We will demonstrate how LLMs can help developing models and configure the solution as well as efficiently guide the business user.

2 - Data-Driven Supply Chain Optimization: Demand Forecasting and Deep Reinforcement Learning

Sohrab Andaz, Amazon, New York, NY, United States

Inventory control is a complex real-world problem that involves handling challenges like seasonal demand, time-varying costs, complex inventory arrival dynamics, and various real-world constraints. In modern supply chain management, effective inventory control requires more than traditional optimization methods. Over the past few years, the Amazon SCOT Demand Forecasting and RL groups have pioneered deep learning models that distill Amazon-scale data into deployable time-series forecasting models and control policies designed to handle these complexities.

In this talk, we will explore both approaches, demonstrating:

1. How deep learning techniques naturally handle supply chain forecasting challenges, including seasonality, cold starts, diverse product categories, and forecast volatility.
2. How deep reinforcement learning can move inventory control beyond the predict-then-optimize framework, allowing practitioners to directly optimize business objectives using historical data.

This talk will also draw on real-world deployment experiences of both deep learning and deep RL policies.

TC26

Summit - 346

Market Dynamics and Financial Modeling

Contributed Session

Chair: Qiang Gao, Baruch College, City University of New York, New York, NY, United States

1 - Augmented Bayesian Estimation for Portfolio Analysis

Yan Wang, The University of Hong Kong, Hong Kong, China, People's Republic of, Peng-Chu Chen

This paper introduces a Bayesian framework for estimating expected asset returns for portfolio selection. This framework relies on a perturbed linear model that predicts out-of-sample portfolio returns, where expected returns are the coefficients to be estimated. The model is fitted to a synthetic dataset via Bayesian regression, and the estimator is computed using a Gibbs sampler, which is both consistent and asymptotically efficient. In a simplified setup, the gap between the out-of-sample mean return of a mean-variance portfolio constructed using this estimator and the best possible out-of-sample mean return is bounded above by the prediction error of the perturbed linear model. As the size of the synthetic dataset increases, the prediction error decreases, narrowing the return gap. This finite sample property ensures that portfolios constructed using our estimator outperform the classical mean-variance, global minimum-variance, and naive portfolios out-of-sample, yielding higher out-of-sample average returns and Sharpe ratios, with or without norm constraints on portfolio weights, as demonstrated by empirical studies.

2 - GameStop in the Lab: Short Squeezes in Experimental Asset Markets

Bradley Paye, Virginia Tech, Blacksburg, VA, United States, Alec Smith, Ross Spoon

This study establishes the existence of explosive short squeezes in laboratory asset markets. The task design includes naturalistic margin and short-sale constraints such that traders can take short and leveraged positions. In most markets, we observe price bubbles in early rounds and the emergence of a small set of traders who take relatively large short positions. Several markets produce outcomes consistent with a short squeeze, featuring explosive price increases that are associated with closing short positions. Following squeezes, prices fall but remain upwardly dislocated from fundamentals for a prolonged period. We also examine how market structure and trader heterogeneity generate squeezes. High cash-to-asset ratios increase the likelihood of large upward price dislocations associated with short squeezes. Our results demonstrate that short-sale constraints can exacerbate, rather than mitigate, asset price bubbles.

3 - Deep hedging of equity linked securities

Jae Wook Song, Hanyang University, Seoul, Korea, Republic of, Namhyoung Kim, Dongwon Ryu

Deep hedging, a sophisticated framework leveraging deep learning for derivative hedging, exhibits commendable efficacy in managing vanilla options. This framework requires the generation of simulated price series for the underlying assets. Equity-linked securities (ELS), as structured financial instruments, derive their returns from the price movements of one or more underlying equities. Consequently, deep hedging applied to ELS requires a time series generation model that captures the correlation among the underlying assets. This study introduces a deep hedging framework designed to effectively manage the risk associated with ELS products featuring intricate pay-off structures. First, a deep learning-based financial time series generation model is employed to generate price series that reflect the asset correlations. Then, a hedger is trained using deep learning to minimize losses for each position, aligned with the specific payoff structure of the ELS. The findings of this study confirm that the proposed approach reduces hedging expenses for ELS compared to the conventional Black-Scholes-Merton framework.

4 - All That Glitters Is Not Gold: The Impact of Certification Costs In Online Labor Markets

Qiang Gao, Baruch College, City University of New York, New York, NY, United States, Jiaru Bai, Paulo Goes, Mingfeng Lin

Third-party skill certification is widely used to address the ubiquitous information asymmetry between workers and employers. A natural way to encourage workers to take certification exams is to make the certification tests free. However, neither theoretical nor empirical studies have examined how employers respond to the decision to offer free tests. This research fills this gap. First, we develop a stylized model to hypothesize the possible effects of zero-cost certification tests on employers' recruitment decisions and their transaction amounts (i.e., contract prices) with workers. Second, we empirically test these hypotheses by exploiting a natural experiment when one of the largest online labor markets unexpectedly removed its certification test fees. Contrary to the platform’s expectations, the offering of free certification tests *reduced* the signaling value of the certifications: Employers’ preference for certified workers decreased (both in terms of hiring likelihood and price paid); more importantly, such preference change appeared economically justified, as workers certified under the new policy are *less* likely to deliver high-quality work. Furthermore, the fall in signaling value is particularly high for inexperienced workers who are still in the process of accruing reviews and are thus more dependent on certifications to secure work. The findings remained consistent over the longer term during our study period. Finally, this study finds evidence that employers have made irrational decisions in hiring after implementing zero-cost certification tests. Our results strongly support the study’s theoretical propositions and have important implications for platforms and practitioners in online labor markets.

TC27

Summit - 347

Sourcing Innovation

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Ersin Korpeoglu, University College London, London, United Kingdom

Co-Chair: Ramazan Kizilyildirim, UCL, LONDON, E14 3SP, United Kingdom

1 - Changing Course: How Analogies Direct Pivots

Jeremy Hutchison-Krupat, University of Cambridge, Cambridge, United Kingdom, Panos Markou

Innovation is exciting because it is full of immense possibility. However, immense possibility is difficult to comprehend. To resolve this, an innovator must interpret or frame the opportunity to reduce the many possible dimensions to just a few prioritized dimensions, which translates to a strategy and a course of action. Yet, rarely does an innovative endeavor follow its original course of action or result in an outcome that replicates what was originally envisioned. Indeed, successful innovation evolves through iterative experimentation to implement a strategy. However, the strategy itself may also require adaptation, otherwise known as a pivot. An effective pivot requires the abandonment of a current strategy, and a decision on a new, more effective, course of action. We conduct two experiments to study how experience with analogous settings enables innovators to address these challenges and pivot effectively. These experiments require individuals to make sense of a difficult problem, adopt and iteratively implement a strategy, and experience progress towards an objective. Then, new information is revealed, which renders the initial strategy ineffective, and requires individuals to pivot their strategy to achieve success. Our results show that analogies are a remarkably effective means to enable individuals to pivot. Moreover, we disentangle the properties of analogies and provide insights on which types of analogies are most likely to prompt a pivot. Our results reveal that while analogies can increase the likelihood to pivot, some types of analogies may lead an innovator to change course, but do so for the wrong reason.

2 - Optimal Feedback in Asymmetric Contests

Hossein Nikpayam, ESMT Berlin, Berlin, Germany

Contests, in which contestants compete at their own expense for prizes offered by a contest holder, have become the foundational primitive of many theories of competition. Recently, the focus in contest research has turned to the role of in-contest performance feedback. The extant literature on feedback has focused on specific ad-hoc policies in symmetric contests and hence failed to more broadly characterize optimal feedback policies. In this paper we solve a general formulation of an asymmetric contest involving feedback, and thus characterize the optimal feedback policy in a very wide class of (stochastic) feedback policies. We find that, in many settings where informative feedback is useful, feedback is optimal when it is both truthful and fully informative.

3 - The Value of Private Feedback in Trial-and-Error Innovation Contests

Zhenyu Hu, National University of Singapore, Singapore, Singapore, Zhi Chen, Yufei Zhu

Firms have increasingly turned to innovation contests as a means of procuring complex industrial innovations from their supplier base. At the start of the contest, the buyer announces the innovation challenge and the award in the form of a valuable supply contract. To address the challenge, it is common that suppliers adopt the trial-and-error approach to develop solutions, and then submit them to the buyer for evaluation as the contest progresses. In some contests, the buyer privately reveals the interim performance of developed solutions to the suppliers during the trial-and-error process (private feedback), but in others, the buyer withholds such information (no feedback). Motivated by such divergent practices, we seek to understand whether providing private feedback (or not) results in higher profits for the buyer. We find that providing private feedback is more profitable for the buyer when either there are many evaluation rounds or when the cost of a trial is high. We further uncover two sources of the value of private feedback: it incentivizes suppliers to conduct more trials ("quantity effect") in a more efficient way ("quality effect"). In contrast, no feedback results in higher profits for the buyer when there are few evaluation rounds and when the cost of a trial is low.

4 - Exclusive OR Not? An Experimental Analysis of Parallel Innovation Contests

RAMAZAN KIZILYILDIRIM, UCL, LONDON, United Kingdom, Cigdem Gizem Korpeoglu, Ersin Korpeoglu, Mirko Kremer

We study parallel innovation contests where organizers seek innovative solutions to a set of problems from independent solvers with limited (financial, time, cognitive) resources. We analyze whether (and when) organizers should discourage solvers from participating in more than one contest. We test contest theory based on a game-theoretic model using controlled laboratory experiments. In the model, a solver's likelihood of winning a contest is determined by the quality of her solution, which improves with her effort and is also influenced by some output uncertainty. Prior theoretical work suggests that organizers should discourage solvers from participating in parallel contests in environments with low output uncertainty, where contest outcomes are primarily driven by solver efforts. In this case, organizers benefit from solvers focusing all of their efforts on a single "exclusive" contest rather than splitting their efforts across multiple "non-exclusive" contests. Our experimental findings depart from theoretical predictions in terms of both solvers' effort choices and the relative profitability of different contest formats. Our main result (and key managerial insight) is that non-exclusive contests are attractive to organizers even in environments with low output uncertainty where theory advocates exclusive contests. We link this result to behavioral tendencies that affect both the average and variability of solvers' efforts in ways that favor non-exclusive contests over exclusive ones.

TC28

Summit - 348

Advanced Modeling and Optimization in Transportation and Manufacturing

Contributed Session

Chair: Xianyue Peng, University of Calfornia, Davis, 4848 El Cemonte Avenue, Davis, CA, United States

1 - Development of artificial neural network-based finite element method surrogate models for predicting material deformation in H-shaped steel intermediate rolling process

Seokkyu Pyo, Sungkyunkwan University, Suwon-si, Korea, Republic of, Hyundeok Jang, Dong-Hee Lee, Sangjin Lee, Hyunseok Jung, Jong-Eun Lee

Structural mechanics simulation for process design in the steel industry relies primarily on a finite element method. The finite element method is traditional and accurate, but the higher the desired resolution, the more time and computational resources are required. The computational resource issue is even more significant in the rolling process, which involves continuous rolling mills. This study proposes an artificial neural network (ANN)-based surrogate model to predict the deformation of H-shaped steel in the intermediate rolling process as an alternative to the finite element method. The finite element simulation data was collected by a fractional factorial design, and features reflecting the relationship between geometry and deformation were added to increase the explanatory power. The trained ANN model successfully predicts the deformation of steel under the untrained process condition, showing a performance of R2 = 0.993 while reducing the prediction time from 6 hours to 5 seconds.

2 - A Strategic Decision-Making Tool for Last Mile Delivery Considering Autonomous Vehicles

Raghavan Srinivasan, BSU, Muncie, IN, United States, Joseph Szmerekovsky

An aggregate delivery capacity plan for last mile delivery services is explored using simulation. This simulation is done considering the use of permanent resources, seasonal resources, crowdsourcing capacity and use of autonomous vehicles or robots to meet the demand. The objective here is to achieve the lowest operational cost in the event of resource attrition and missed deliveries. The results from the simulation show that autonomous vehicles can reduce dependence on seasonal capacity. This use of automation can both decrease cost and increase service level. Also, the operational challenges associated with the use of automation for last mile delivery aggregate capacity planning are discussed.

3 - Application of Agent-Based Simulation Model in Demand Estimation of CNG Stations

Yinglei Li, Breakthrough - a USVenture company, BELLEVUE, WA, United States, Andy Martinelli

An energy company attempts to leverage the information of shipment volume and routes to better estimate the current and potential demand their station locations could possibly generate. To achieve this goal, an agent-based simulation model has been developed to quantify the potential fuel demand of each CNG station based on the flow of network loads. In the simulation model, each agent represents a driver. Each agent can be customized by the users to specify the driver’s habits and preference in terms of eating, parking for resting, bathroom, and brand requirement during the tours. Different probability functions can be assigned to the agent to determine when and where an agent would stop and re-fuel. The results can help the energy company with the following: generating different agent scenarios, evaluating consumption estimates at each CNG station, partnering with transportation providers and shippers who utilize the stations, and planning for new complimentary locations.

4 - Optimal Lock Maintenance of the U.S. Inland Waterway Transportation System under Uncertainty Using Deep Reinforcement Learning

Maryam Aghamohammadghasem, University of Arkansas, Fayetteville, AR, United States, Jose Azucena, Haitao Liao, Shengfan Zhang

The interconnected infrastructure, including the lock and dam systems, is essential to the inland waterway transportation system (IWTS) 's continued operation. However, as the IWTS ages, more unforeseen events, such as failure, occur, which results in major unplanned maintenance, transportation delays, and financial losses.

A Python-enhanced NetLogo simulation tool is developed to assess the effects of lock-induced disruptions on IWTS. Deep reinforcement learning is used to determine the best maintenance plans that maximize cargo throughput on the IWTS. The locks are grouped together using a machine-learning technique, and individual repair teams are assigned to each cluster. If the corresponding repair teams for every individual cluster are overworked, outsourcing to other idle groups with higher expenses may be an option. Additionally, a maintenance group consists of a variety of individuals with varying specialties who are assigned to a particular failure requiring specialized knowledge. The repair crew's travel time is factored in to make it seem more realistic.
The potential of the developed simulation and machine learning-based approach for IWTS maintenance optimization is demonstrated through a case study of the lower Mississippi River system and the McClellan-Kerr Arkansas River Navigation System.

5 - A multi-scale perimeter control and route guidance system for large-scale road networks

Xianyue Peng, University of California-Davis, Davis, CA, United States, Hao Wang, Shenyang Chen, Michael Zhang

Perimeter control and route guidance are effective ways to reduce traffic congestion and improve traffic efficiency by controlling the spatial and temporal traffic distribution on the network. This study presents a multi-scale perimeter control and route guidance framework for controlling traffic in large-scale networks. The network is first partitioned into several subnetworks (regions) with traffic in each region governed by its macroscopic fundamental diagram (MFD), which forms the macroscale network (upper level). Each subnetwork, comprised of actual road links and signalized intersections, forms the microscale network (lower level). At the upper level, a joint perimeter control and route guidance model solves the region-based inflow rate and hyper-path flows to control the accumulation of each region and thus maximize the throughput of each region. At the lower level, a perimeter control strategy integrated with a backpressure policy determines the optimal signal phases of the intersections at the regional boundary. At the same time, a route choice model for vehicles is constructed to meet hyper-path flows and ensure the intra-region homogeneity of traffic density. The case study results demonstrate that the proposed joint control outperforms either perimeter control or backpressure control in regulating regional accumulation, thereby achieving higher throughput.

TC29

Summit - 420

Sequential Decision Making

Invited Session

OPT: Integer and Discrete Optimization

Chair: Rohan Ghuge, Georgia Institute of Technology, Atlanta, United States

Co-Chair: Yihua Xu, Georgia Institute of Technology, Atlanta, GA, United States

1 - Exponential Pricing Algorithms for Online Resource Allocation

Rohan Ghuge, Georgia Institute of Technology, Atlanta, GA, United States, Yifan Wang, Sahil Singla

Online resource allocation problem is a central problem in many areas of Computer Science, Operations Research, Economics, and Networks. In this problem, requests arrive sequentially, and each request can be satisfied in multiple ways, each consuming a certain amount of resources while generating different values. The objective is to maximize the overall value without exceeding the {budget} for any resource. Of particular interest in this problem is the design of \emph{pricing} algorithms that present each incoming request with prices for the resources, enabling the request to make a value-maximizing decision based on current resource prices. The goal is to achieve a $(1+\epsilon)$-approximation to the hindsight optimum, where $\epsilon>0$ is a small constant.

2 - Stochastic Scheduling: Strategies for Abandonment Management

Yihua Xu, Rice University, Houston, TX, United States, Rohan Ghuge, Sebastian Perez-Salazar

We explore a job scheduling model driven by applications with impatience, such as in call centers and cloud computing. We have a single server and n jobs with known non-negative values. These jobs also have unknown stochastic service and departure times with known independent distributional information. We aim to design a policy that maximizes the expected total value obtained from jobs run on the server. Natural dynamic programming formulations face the curse of dimensionality. Furthermore, our problem is demonstrably NP-hard. Hence, we focus on policies that can provide good approximation factors compared to the optimal value. We provide a compact linear program relaxation that upper bounds the performance of the optimal policy and examine the quality of this LP relaxation by bounding its integrality gap by constants. As a byproduct of our analysis and mild assumptions on the service times supports, we provide a polynomial-time computable policy that is a (1/2)(1 − 1/e) approximation of the optimal expected reward. Our methodology is flexible, allowing additional constraints to be incorporated. We develop efficient approximation algorithms with provable guarantees for extensions like job release times, deadlines, and knapsack constraints. Additionally, we consider jobs having independent and identically distributed service times and show that the natural greedy policy guarantees half of the optimal expected value. We provide empirical validation of our policies via numerical experiments.

3 - New guarantees for online selection over time

Sebastian Perez Salazar, Rice University, Houston, TX, United States

We revisit the prophet inequality over time problem. In this problem, the decision-maker can choose how long to select a value instead of stopping with only one realized value, as in the classical setting. We provide improved bounds based on density arguments and convex optimization.

4 - Scheduling Flexible Jobs with Predicted Lengths Non-preemptively for Minimum Span

Mozhengfu Liu, Northwestern University, Evanston, IL, United States, Xueyan Tang

We study an online model in which a sequence of so-called flexible jobs, each with a starting deadline and a processing length, are released over time. The jobs are to be started by an online scheduler no later than their starting deadlines and will run non-preemptively. Non-preemption means that once the job is scheduled to start, it is not allowed to be paused and resumed later. There is no limit on the number of simultaneously running jobs. The objective is to minimize the span, i.e., the time duration for which at least one job is running. We investigate the problem in the learning-augmented setting: when each job is released, a predicted length rather than the real length of it is revealed. We show asymptotically tight deterministic bounds for its competitiveness. Minimizing the span (sometimes called active time) is foundational to solving busy time problem. The busy time problem, in which flexible jobs are to be scheduled into machines to minimize the accumulated running time of all machines, is closely related to optimizing the monetary costs of renting machines in cloud computing and saving the power usage of machines in energy-efficient computing.

TC30

Summit - 421

Advances in Optimization for Machine Learning

Invited Session

OPT: Nonlinear Optimization

Chair: Pratik Rathore, 94305

Co-Chair: Zachary Frangella, Stanford University, Stanford, CA, United States

Co-Chair: Madeleine Udell, Stanford University, Stanford, CA, 94305, United States

1 - Convex Optimization Reformulations for Deep Neural Networks

Mert Pilanci, Stanford University, Stanford, CA, United States

In this talk, we introduce an analysis of deep neural networks through convex optimization. We begin by introducing exact convex optimization formulations for ReLU neural networks. This approach demonstrates that deep networks can be globally trained through convex programs, offering a globally optimal solution. Our results further establish an equivalent characterization of neural networks as high-dimensional convex Lasso models. These models employ a discrete set of wedge product features, and apply sparsity-inducing convex regularization to fit data. This framework provides an intuitive geometric interpretation where the optimal neurons represent signed volumes of parallelotopes formed by data vectors. We introduce numerical optimization schemes to obtain near-optimal solutions efficiently.

2 - Online Learning Guided Quasi-Newton Methods: Improved Global Non-Asymptotic Guarantees

Ruichen Jiang, University of Texas at Austin, Austin, TX, United States, Qiujiang Jin, Aryan Mokhtari

Quasi-Newton (QN) methods are popular iterative algorithms known for their superior practical performance compared to Gradient Descent (GD)-type methods. However, the existing theoretical results for this class of algorithms do not sufficiently justify their advantage over GD-type methods. In this talk, we discuss our recent efforts to address this issue. Specifically, in the strongly convex setting, we propose the first “globally” convergent QN method that achieves an explicit “non-asymptotic superlinear” rate. We show that the rate presented for our method is provably faster than GD after at most $O(d)$ iterations, where d is the problem dimension. Additionally, in the convex setting, we present an accelerated variant of our proposed method that provably outperforms the accelerated gradient method and converges at a rate of $O(\min\{1/k^2, \sqrt{d \log k }/ k^{2.5}\})$, where k is the number of iterations. To attain these results, we diverge from conventional approaches and construct our QN methods based on the Hybrid Proximal Extragradient (HPE) framework and its accelerated variants. Furthermore, a pivotal algorithmic concept underpinning our methodologies is an online learning framework for updating the Hessian approximation matrices. Specifically, we relate our method's convergence rate to the regret of a specific online convex optimization problem in the matrix space and choose the sequence of Hessian approximation matrices to minimize its overall regret.

3 - Schedule-Free Learning

Aaron Defazio, Meta Platforms, Inc., New York, NY, United States

Existing learning rate schedules that do not require specification of the optimization stopping time T are greatly out-performed by learning rate schedules that depend on T. We propose an approach that avoids the need for this stopping time, exhibiting state-of-the-art performance compared to standard schedules across a wide range of problems, including both convex problems and large-scale deep learning problems. Our method introduces no additional hyper-parameters over standard SGD with momentum. Our method is a direct consequence of a new theory we develop that unifies scheduling and iterate averaging. An open source implementation of our method is available.

4 - Policy Optimization with Compatible Mirror Approximation

Lin Xiao, FAIR @ Meta, Seattle, WA, United States, Zhihan Xiong, Maryam Fazel

We propose Compatible Mirror Policy Optimization (CoMPO), a framework that incorporates general function approximation into policy mirror descent methods for reinforcement learning. In contrast to the popular approach of using the squared L2 norm to measure function approximation errors, CoMPO uses the Bregman divergence induced by the same mirror map as in policy projection. Such a compatibility bridges the gap between theory and practice: not only does it lead to fast linear convergence with general function approximation, but it also includes several well-known practical reinforcement learning methods as special cases, including Soft Actor-Critic, immediately providing them with strong convergence guarantees.

5 - A Non-stochastic Control Approach to Optimization

Xinyi Chen, Princeton University, Princeton, NJ, United States, Elad Hazan

Selecting the best hyperparameters for a particular optimization instance, such as the learning rate and momentum, is an important but nonconvex problem. As a result, iterative optimization methods such as hypergradient descent lack global optimality guarantees in general.We propose an online nonstochastic control methodology for mathematical optimization. First, we formalize the setting of meta-optimization, an online learning formulation of learning the best optimization algorithm from a class of methods. The meta-optimization problem over gradient-based methods can be framed as a feedback control problem over the choice of hyperparameters, including the learning rate, momentum, and the preconditioner. We show how recent methods from online nonstochastic control can be applied to develop a convex relaxation, and obtain regret guarantees vs. the best offline solution. This guarantees that in meta-optimization, given a sequence of optimization problems, we can learn a method with performance comparable to that of the best method in hindsight from a class of methods. We end with experiments on a variety of tasks, from regression to deep neural network training, that demonstrate the practical effectiveness of our method.

TC31

Summit - 422

Recent Advances of Interior Point Methods for (Non)linear Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Madeleine Udell, Stanford University, Stanford, CA, 94305, United States

Co-Chair: Wenzhi Gao, Stanford University, Stanford, CA, 94305, United States

1 - When Does Primal Interior Point Method Beat Primal-Dual?

Wenzhi Gao, Stanford University, Stanford, CA, United States, Huikang Liu, Madeleine Udell, Yinyu Ye

It is widely accepted that primal-dual interior point method (IPM) is the most efficient IPM variant for linear optimization. In this paper, we show this is not always the case by inspecting IPM in the pure primal form. We show that stability of the primal scaling matrix enables several computational techniques for acceleration, including efficient preconditioning with iterative methods, approximate search direction generation and robustness to perturbation.

2 - Clarabel: An Interior Point Solver for Conic Optimization

Yuwen Chen, University of Oxford, Oxford, United Kingdom, Paul Goulart

The homogeneous self-dual embedding (HSDE) is a popular technique that unifies optimality and infeasibility detection of a convex problem and has been widely used in convex interior-point solver. However, it only allows a linear cost in the objective and we need to eliminate the quadratic term in the objective function, replacing it with an epigraphical upper bound and an additional second-order cone constraint in the objective.

We presented the interior-point solver, Clarabel, for conic optimization with quadratic objectives by using the homogeneous embedding for infeasibility detection, which tackled the issue of transforming QPs into SOCPs. The solver is fully written in both Julia and Rust languages. We support second-order cones, exponential cones, power cones and positive semidefinite cones in Clarabel. Our solver is competitive with the cutting-edge conic solvers on various benchmarks and outperforms them in terms of time and numerical stability on QPs and SOCPs.

3 - The Nonsymmetric Conic Optimizer in Copt

Joachim Dahl, Cardinal Operations, Copenhagen, Denmark

Over the last two decades there has been much research into interior-point algorithms for conic optimization, especially linear programming, second-order cone and semidefinite optimization, which has been implemented in several commercial and research packages.

In comparison, extensions to other (non-symmetric) cones are much less studies. In this talk we consider one such extension, namely the exponential cone, which has recently been added to the COPT conic optimizer.

4 - Polynomial-TIME Interior-Point Methods for Convex Optimization Beyond Symmetric Cones

Mehdi Karimi, Illinois State University, Normal, IL, United States, Levent Tuncel

Since Nesterov and Nemirovski’s seminal book in 1994, polynomial-time interior-point (IP) methods for convex optimization have been extensively investigated. While the focus for creating solvers has been on optimization over symmetric cones, several recent efforts have aimed to develop robust solvers for optimization over other interesting convex sets, such as the exponential cone and the quantum relative entropy (QRE) cone. This presentation introduces our Domain-Driven IP algorithms and the corresponding solver DDS. Our approach and the solver DDS accept every convex set represented by a computationally efficient self-concordant barrier. We present the recent developments in our approach and software DDS. In addition to accepting problems with several function/set constraints, the newest version, DDS 2.2, has modifications for solving QRE programming problems, potentially combined with many other constraints. Concluding the presentation, we showcase numerical results demonstrating the efficacy of DDS in solving diverse problem classes.

5 - Recent Advances in Quantum Interior Point Methods

Tamás Terlaky, Lehigh University, Bethlehem, PA, PA, United States, Mohammadhossein Mohammadisiahroudi, Zeguan Wu, Pouya Sampourmahani

Various Quantum Interior Point Methods (QIPMs) were proposed and analyzed in the past years for Conic Linear Optimization (CLO) problems. Those include: Inexact Infeasible, Inexact Feasible QIPMs and Quantum Central Path algorithms. Some QIPMs are enhanced with preconditioning and iterative refinement methodologies. Novel quantum variants inspired novel classic variants too.

In this talk, at a high level, we review the key features of the various QIPMs and their classic counterparts, and highlight the advantages and challenges offered by quantum computing, the pros and cons of the algorithm paradigms.

TC32

Summit - 423

Reformulations and Solution Approaches for Mixed-Integer Nonlinear Optimization

Invited Session

OPT: Global Optimization

Chair: Andres Gomez, University of Southern California, Los Angeles, CA, United States

1 - Real-TIME Solution of Mixed-Integer Quadratic Programs Using Decision Diagrams

Shaoning Han, National University of Singapore, Singapore, Singapore, Andres Gomez, Leonardo Lozano

We consider mixed-integer quadratic optimization problems with banded matrices and indicator variables. These problems arise pervasively in statistical inference problems with time-series data, where the banded matrix captures the temporal relationship of the underlying process. In particular, the problem studied arises in monitoring problems, where the decision-maker wants to detect changes or anomalies. We propose to solve these problems using decision diagrams. In particular we show how to exploit the temporal dependencies to construct diagrams with size polynomial in the number of decision variables. We also describe how to construct the convex hull of the set under study from the decision diagrams, and how to deploy the method online to solve the problems in milliseconds via a shortest path algorithm.

2 - Isotonic Optimization with Markup Costs

Shuai Li, Georgia Institute of Technology, Atlanta, GA, United States, Xin Chen, Weijun Xie

This paper addresses a new mixed-integer nonlinear optimization problem that extends the generalized isotonic optimization problem by incorporating markup cost whenever a change occurs. This formulation considers a directed path graph and puts isotonic constraints on each arc. The costs incurred at each node consist of a markup cost and a convex cost, and our objective is to minimize the overall costs. The formulation covers various applications, such as reduced isotonic regression, stochastic lot-sizing, joint inventory control, and pricing with costly price markdown. After proving an optimality condition, we introduce a shortest path algorithm achieving $O(n^2 \log n)$ complexity to solve this problem to optimality. Besides, after assuming the minima of those convex functions follow a non-decreasing order, we can improve computational efficiency to $O(n)$. We then extend this model by making the graph from a directed path to an arborescence and show that the current shortest path algorithm fails generally.

3 - New Finite Hierarchy for Disjoint Bilinear Programs

Mohit Tawarmalani, Purdue University, West Lafayette, IN, United States

We introduce new hierarchy for disjoint bilinear programs. These hierarchies utilize rational functions as product-factors and improve upon relaxations obtained using disjunctive programming and reformulation-linearization technique providing new insights relating these hierarchies with one another. Our hierarchies have geometric and algebraic underpinnings and converge at a finite level. Finally, we leverage our techniques to develop convex relaxations for concave-convex functions and new hierarchies for facial disjunctive programs.

4 - A Convexification-Based Outer-Approximation Method for Convex and Nonconvex MINLP

Zedong Peng, Purdue University, West Lafayette, IN, United States, Kaiyu Cao, Kevin Furman, Can Li, Ignacio Grossmann, David Bernal Neira

The advancement of domain reduction techniques has significantly enhanced the performance of solvers in mathematical programming. This work delves into the impact of integrating convexification and domain reduction techniques within the Outer- Approximation method. We propose a refined convexification-based Outer-Approximation method alongside a Branch-and-Bound method for both convex and nonconvex Mixed-Integer Nonlinear Programming problems. These methods have been developed and incorporated into the open-source Mixed-Integer Nonlinear Decomposition Toolbox for Pyomo-MindtPy. Comprehensive benchmark tests were conducted, validating the effectiveness and reliability of our proposed algorithms. These tests highlight the improvements achieved by incorporating convexification and domain reduction techniques into the Outer-Approximation and Branch-and-Bound methods.

5 - A Polyhedral Study on L-Natural-Convex Minimization and Its Mixed-Integer Extension

Qimeng Yu, Université de Montréal, Montréal, QC, Canada, Simge Kucukyavuz

L-natural-convex functions are a class of nonlinear functions defined over integral domains. Such functions are not necessarily convex, but they display a discrete analogue of convexity. In this work, we explore the polyhedral structure of the epigraph of any L-natural-convex function and provide a class of valid inequalities. We show that these inequalities are sufficient to describe the epigraph convex hull completely, and we give an exact separation algorithm. We further examine a mixed-integer extension of this class of minimization problems and propose strong valid inequalities. We establish the connection between our results and the valid inequalities for some structured mixed-integer sets in the literature.

TC33

Summit - 424

Stochastic Programming with Endogenous Uncertainty.

Invited Session

OPT: Optimization Under Uncertainty

Chair: Samuel Affar, University of Tennessee, Knoxville, Knoxville, TN, United States

1 - A Successive Refinement for Solving Stochastic Programs with Decision-Dependent Random Capacities

Samuel Affar, University of Tennessee, Knoxville, Knoxville, TN, United States, Hugh Medal

We study a class of two-stage stochastic programs in which the second stage includes a set of components with uncertain capacity, and the expression for the distribution function of the uncertain capacity includes first-stage variables. Thus, this class of problems has the characteristics of a stochastic program with decision-dependent uncertainty. A natural way to formulate this class of problems is to enumerate the scenarios and express the probability of each scenario as a product of the first-stage decision variables; unfortunately, this formulation results in an intractable model with a large number of variable products with high-degree. After identifying structural results related to upper and lower. bounds and how to improve these bounds, we present a successive refinement algorithm that successively and dynamically tightens these bounds. Implementing the algorithm within a branch-and-cut method, we report the results of computational experiments that indicate that the successive refinement algorithm significantly outperforms a benchmark approach. Specifically, results show that the algorithm finds an optimal solution before the refined state space becomes too large.

2 - Searching Model with Decision-Dependent Uncertainty

Luis Novoa, James Madison University, Harrisonburg, VA, United States

We propose a decision-dependent uncertainty stochastic programming model, along with reformulations and tailored decomposition solution methods, aimed at prescribing optimal search and patrolling strategies to locate a target that follows an uncertain path across discrete time periods. We assume that the target's path selection is contingent on proposed surveillance strategies—number and positioning of mobile searchers and static sentries with diverse capabilities— consequently making uncertainty modeling decision-dependent. This dependency is addressed through the definition of a coupling function that describes how the probability distribution associated with the random target's path is influenced by surveillance decisions. The problem at hand results in a large-scale nonconvex MIP, which can be reformulated as a MILP requiring tailored solution approaches. We propose an efficient branch-and-Benders-cut single-tree solution algorithm that leverages the unique structure of our problem.

3 - Informative Path Planning with Limited Adaptivity

Rayen Tan, University of Michigan, Ann Arbor, MI, United States, Rohan Ghuge, Viswanath Nagarajan

We consider the informative path planning (IPP) problem in which a robot interacts with an uncertain environment and gathers information by visiting locations. The goal is to minimize its expected travel cost to cover a given submodular function. Adaptive solutions, where the robot incorporates all available information to select the next location to visit, achieve the best objective. However, such a solution is resource-intensive as it entails recomputing after every visited location. A more practical approach is to design solutions with a small number of adaptive "rounds", where the robot recomputes only once at the start of each round. In this paper, we design an algorithm for IPP parameterized by the number k of adaptive rounds, and prove a smooth trade-off between k and the solution quality (relative to fully adaptive solutions). We validate our theoretical results by experiments on a real road network, where we observe that a few rounds of adaptivity suffice to obtain solutions of cost almost as good as fully-adaptive ones.

4 - Probing-enhanced stochastic programming

Jim Luedtke, University of Wisconsin-Madison, Madison, WI, United States, Jeff Linderoth, Zhichao Ma

We consider a two-stage stochastic program where the decision-maker has the opportunity to obtain information about the distribution of the random variables X through a set of discrete actions that we refer to as probing. Probing allows the decision-maker to observe components of a random vector Y that is jointly-distributed with X. We propose a three-stage optimization model for this problem, where the first-stage variables select components of Y to observe. In the case that X and Y have finite support, a model of Goel and Grossmann can be applied to obtain a formulation of this problem whose size is proportional to the square of cardinality of the sample space of the random variables. We propose to solve the model using bounds obtained from an information-based relaxation, combined with a branching scheme that enforces the consistency of decisions with observed information. The branch-and-bound approach can naturally be combined with sampling in order to estimate both lower and upper bounds on the optimal solution value even for problems with continuous distribution. We demonstrate the approach on instances of a stochastic facility location problem.

TC34

Summit - 425

Advancements in Urban Transit Systems

Contributed Session

Chair: Bolong Zhou, The Hong Kong University of Science and Technology, Hong Kong, N/A, Hong Kong

1 - Optimizing the schedule of UAV-assisted urban subway inspection services

Bolong ZHOU, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Wei Liu, Hai Yang

The periodic inspection and maintenance of subway facilities are essential for ensuring passenger safety. However, the current manual inspection practices conducted by expert engineers are time-consuming, costly, and pose risks to workers. Unmanned aerial vehicles (UAVs) offer a promising solution for automatically inspecting subway facilities. This paper investigates an operational-level synchronized optimization problem, aiming to determine an optimal inspection timetable while simultaneously optimizing working schedules for both human teams and UAVs. Demand heterogeneity is taken into account since the variety of facilities and equipment in subway tunnels may have different required inspection cycles. A compact Mixed-Integer Linear Program (MILP) model is developed to solve this NP-hard problem. We propose an enumeration-based reformulation and develop an exact algorithm that combines Benders decomposition and Dantzig-Wolfe decomposition within a branch-and-price framework to solve the MILP model efficiently. The approach is strengthened by implementing several acceleration strategies. Extensive numerical experiments have been carried out. The results show that our proposed optimization model and algorithms can find the optimal solution for real-world scale instances, resulting in cost savings and improved efficiency. Furthermore, we highlight the benefits of integrated optimization by comparing our solution approach with a sequential method that addresses inspection timetables and working schedules separately.

2 - Mobility, Equity, and Economic Impact of Transit in the Chicago Metropolitan Region

Omer Verbas, Argonne National Laboratory, Lemont, IL, United States, Taner Cokyasar, Joshua Auld, Aymeric Rousseau, Seamus Joyce-Johnson, Jim Aloisi, Anson Stewart, Gabriel Barrett

Transit has a low modal share in most US metropolitan regions due to high car ownership and sparse, suburban land use outside the urban core. However, it is still the backbone of urban travel. In the Chicago Metropolitan Region, the regional modal share of transit was 5. 5% before the pandemic, and it is expected to be around 3.1% in 2025 due to increased negative perception of transit and shift in telecommuting patterns. This study focuses on a complete transit removal scenario in Chicago metropolitan region resulting in a removal of ~25K scheduled service trips and 1.35M single-leg riders. The scenario also envisions a 30% increase in car ownership, where households without cars buy a car, and households with a single car buy an additional one. Using POLARIS, an agent-based travel demand modeling and multimodal traffic simulation platform, both scenarios – 2025 baseline and transit removal – are simulated. Compared to baseline, average speeds decrease by 16% in the region and 33% in the city. Similarly, average travel times increase by 14% in the region and 35% in the city. In the entire region, 12% (1.6M) of non-work activities and 9% (1.9M) of overall activities get cancelled. In the city, 27% (1.0M) non-work activities, and 20% (1.2M) of overall activities get cancelled. The activity cancellations are suffered disproportionately by women and low-income groups.

3 - Evaluation of Travel Demand Models for On-demand Multimodal Transit Systems

Ryan Rodriguez, Georgia Institute of Technology, Atlanta, GA, United States, Hongzhao Guan, Pascal Van Hentenryck

Travel demand modeling is important for the design of transit systems because it allows planners to identify and analyze the markets and produces ridership estimates. This presentation introduces a travel demand modeling study for a specific type of transit systems: On-Demand Multimodal Transit Systems (ODMTS). ODMTS are an emerging transit concept that seamlessly integrates high-frequency fixed-routes like rail and bus rapid transit with on-demand shuttles into a single synchronized system, addressing the first/last mile problem faced by transit agencies. In recent years, ODMTS have been deployed in two cities in Georgia: Atlanta (2022) and Savannah (2024). With information collected from these real-world deployments, this study combines data from multiple sources such as demographic, geographic, and historical trip data to generate estimations at the census block level. Machine learning frameworks are employed to classify the demand of each census block into the following categories: empty, low, medium, or high. A detailed analysis between intra-dataset evaluation and inter-dataset evaluation has provided valuable insights into understanding the travel behavior of ODMTS riders. Specifically, intra-dataset evaluation studies the travel behavior within specific urban areas, serving as a basis for developing a complicated mode choice model---a critical piece overlooked in a few previous ODMTS studies. Moreover, through inter-dataset evaluation, this study reveals the machine learning framework's potential effectiveness in transferring demand knowledge across different urban areas, which is crucial for informing the design and development of future ODMTS programs.

4 - Designing Incentive Bundles for Multimodal Transportation Systems

Jason Lu, University of Michigan, Ann Arbor, MI, United States, Neda Masoud

In Mobility-as-a-Service (MaaS) systems, the incorporation of multimodal transport, such as transit and ridesharing, has the potential to provide greater accessibility, operational efficiency, and user convenience of mobility services. To encourage effective system utilization and improve social equity in a multimodal setting, incentives such as subsidizing costs for users could be developed. In this paper, we propose a novel incentive scheme that bundles subsidies across multiple modes for a MaaS system under a fixed budget. Bundles are designed to encourage multimodal usage of the system with the goal of enhancing overall social welfare across all system users and improving equity within the system. An optimization framework is developed to find an optimal incentive strategy that subsidizes user costs across combinations of multiple modes. Numerical results on a multimodal network demonstrate that incentive bundles can significantly improve social welfare and system efficiency compared to traditional MaaS Systems while also bridging the equity gap among users.

5 - PICASA - Optimal Capacity allocation in Supply chain for better speed offering

V N Sai Anjani Kumar Kudupudi, Flipkart India Private Limited, Bengaluru, India, Aditya Siva Sai Prasanth Bandaru, Gowtham Bellala, Vikas Goel, Anjan Goswami, Rishabh Bhat, Sarvesh Desai

Delivery Speed is one of the critical metrics for an e-commerce company. Fulfillment centers (FC) of an e-tailer typically assign approved orders to waves (defined by some cut-off timings) and dispatch the shipments to the next destination in the supply chain. During periods of high demand (e.g., a sale event), orders end up getting assigned to future waves leading to spillage in the supply chain impacting the delivery speed. Our goal is to design a system that will ensure a consistent delivery speed and experience to key buying cohorts. A buying cohort corresponds to a combination of customer segment (e.g., customers enrolled in loyalty programs) and product segment (e.g, speed sensitive products). Spillage protection to key buying cohorts is achieved by reserving dedicated capacity at various assets/nodes in the supply chain. However, a capacity segmentation strategy can lead to inefficient utilization of these capacities. As a result, it becomes imperative that we optimize for trade-offs between spillage protection and capacity under-utilization to provide better speed. This problem is especially challenging during sale events when the demand patterns are highly unpredictable. In this work, we developed a stochastic optimization model which dynamically determines the optimal allocation of capacity across the cohorts with an objective of providing better protection to key buying cohorts under minimal under-utilization. This model has been deployed in major sales and helped us to offer 1600 bps differential 2-day delivery speed for the high priority cohort.

TC35

Summit - 427

Middle Mile Optimization and Service Network Design

Invited Session

Transportation Science and Logistics (TSL)

Chair: Sarah Powell, University of Iowa, Iowa City, IA, United States

1 - Underground Freight Transportation for Middle Mile Delivery

Sarah Powell, University of Iowa, Iowa City, IA, United States, Ann Campbell, Mojtaba Hosseini

The use of tunnels for package delivery is receiving increased attention as it provides an efficient and environmentally friendly alternative to truck delivery. Tunnel delivery removes trucks from the road decreasing traffic, emissions, and noise pollution, and provides a more reliable method of delivery in cases of extreme weather. We consider the use of tunnels for middle mile delivery and evaluate the impact on the number of orders that can arrive at neighborhood distribution centers in time for next-day delivery, which is a priority in e-commerce. We provide an integer programming model for this new problem and consider different options for combining and sorting deliveries. We determine the tunnel capacity needed to meet all next-day delivery requirements. We provide insights on optimal dispatch schedules for tunnel vehicles when tunnel capacity is binding. In a set of computational tests, we compare the use of tunnels to truck delivery in the middle mile and determine the conditions under which using tunnels allows higher levels of next-day delivery.

2 - Valid Inequalities for Service Network Design Models

Lacy Greening, Arizona State University, Tempe, AZ, United States, Santanu Dey, Alan Erera

This presentation will discuss two new classes of valid inequalities for service network design models (applicable to both arc- and path-based formulations). Results from a computational study using both industry-based and the so-called CANAD instances will demonstrate their effectiveness to strengthen the linear programming (LP) relaxations.

3 - Dynamic Service Network Design

Alexander Bode, TU Braunschweig, Braunschweig, Germany, Mike Hewitt, Dirk Mattfeld, Marlin Ulmer

Companies in the LTL industry have a network of transshipment locations and transportation services between them. Generally, transportation services are scheduled tactically to deliver the commodities being shipped. However, it is possible that the demand for transportation services exceeds the supply. Commodities that cannot be delivered with the regular transportation services are called excess cargo. Our problem deals with the issue of how companies can transport excess cargo as cost-efficiently as possible with the help of third-party providers. Excess cargo cannot be predicted and occurs randomly. Decisions must therefore be made dynamically in order to be able to react to the new commodities that arise. We therefore do not consider the entire time horizon but use a rolling horizon approach to decompose the problem over time and solve it in individual steps. The problem is modeled using a MIP and a sequential decision process. Intermediate destination nodes are introduced as a concept. The solution method is based on this concept and a two-part Cost Function Approximation (CFA) is developed, which can estimate the costs of decisions as accurately as possible. The first part of the CFA estimates the near future by estimating the consolidation possibilities of the known commodities. The second part of the CFA estimates the costs in the distant future and uses a measure based on slack to evaluate the progress of the commodities. We test our CFA on real data and the results clearly show the effectiveness of the method.

4 - Improvement Heuristics for Large Scale Ltl-Service Network Design

Ravit Pichayavet, Georgia Tech ISyE, Atlanta, GA, United States, Daniel Ulch, Alejandro Toriello, Alan Erera

This work presents heuristic solution approaches that solve an extra large-scale less-than-truckload service network design problem. The problem comes primarily from the strategic flow and load planning of our industry partner, one of the biggest parcel delivery companies in the US. The problem seeks to jointly decide the flow and load planning, where the flow plan specifies the sequence of sorting facilities (location) for package flow, and the load plan determines a detailed timing for a truck dispatch (load) that carries packages to their destinations. We introduce a marginal cost path algorithm, a local search improvement heuristic that iteratively selects a local neighborhood and re-paths flow commodity when marginal cost saving exists. We benchmark the algorithm with slope-scaling heuristics, a well-known heuristic approach that can be applied up to the size of our instance. Two computational experiments motivated by the practical use cases of our industry partner, including primary-flow plan improvement that seeks to find adjustments to the company's existing plan and contingency response plan that seeks to find cost-efficient adjustments in response to the capacity disruption at the sorting facility, are conducted.

TC36

Summit - 428

Crowdsourced Logistics Operations

Invited Session

TSL: Freight Transportation

Chair: Ali Sabzevari Zadeh, University of Alabama, Tuscaloosa, AL, United States

1 - Crowdshipping Problem with Dynamic Compensations and Transshipment

Barış Yıldız, Koç University, Istanbul, Turkey, Ali Sardag, Kerim U. Kizil

Rapid urban growth and consequent increase in e-commerce demand make urban logistics a harder task than ever. The growing size of urban delivery operations not only entails operational challenges but also generates several negative externalities, such as increased traffic, pollution, noise, and accidents. This trend creates a pressing need for efficient delivery mechanisms that are more economical and environmentally friendly than existing systems. Crowdshipping, wherein ordinary members of the society partake in delivery operations for a small compensation, is one of the answers that cater to this need and has attracted considerable research interest recently. However, designing compensation mechanisms to prompt efficient participation from the public remains largely unexplored in the literature. In this study, we devise a dynamic compensation scheme for crowdshipping operations in a many-to-many express delivery framework, where the crowdshipper compensations are determined based on spatial and temporal distributions of the delivery demand and continually updated during the service time to leverage the crowd participation as needed. To address the resulting complex network management problem, we derive analytical solutions for compensation optimization and use these results along with effective pruning strategies to build a lookup table to simultaneously determine package routes and compensation offers in real time. Computational studies and extensive simulations conducted with real-world data show that our proposed approach can provide significant cost savings and considerably reduce operational costs and other transport-related negative externalities when compared to classical delivery modes, crowdshipping with static compensations, and crowdshipping without transshipment.

2 - Query-based Request Reallocation for Collaborative Urban Transportation

Steffen Elting, University of Vienna, Vienna, Austria, Jan Fabian Ehmke, Margaretha Gansterer

To reduce costs and greenhouse gas emissions in the last-mile delivery sector, multiple transportation service providers can engage in an auction-based horizontal collaboration. With the help of a central auctioneer, delivery orders are reallocated among the carriers so that the total driving duration can be decreased compared to isolated planning. To minimize information sharing, the auctioneer conducts a combinatorial auction which usually requires an exponential number of bids to be communicated where each bid is equivalent to solving a complex vehicle routing problem. We investigate an integrated, query-driven approach to limit the number of bids while still obtaining an efficient order allocation. Rather than executing the bundling and bidding steps of the auction separately, the auctioneer iteratively poses a few value queries to the carriers at a time and uses their responses to train regression models that are used as surrogate fitness functions when searching for the next queries. We simulate a collaboration in the city of Vienna, Austria, with three carriers and 50 customers each. Our results show that the proposed approach can be up to 10% better than the sequential benchmark from the literature and up to 20% better than a random benchmark. However, these results depend on the number of value queries that the auctioneer is allowed to ask, as the regression-based approaches require sufficient training data to work well.

3 - Would you deliver your neighbor's package? Crowdsourced last-mile delivery

Ali Sabzevari Zadeh, University of Alabama, Tuscaloosa, AL, United States, Iman Dayarian

We study the problem of last-mile delivery in the context of an e-retailer offering two types of services: same-day delivery through parcel lockers (PL), and next-day home delivery (HD). Daily, PL customers place their orders to pick up from their preferred parcel locker and period, among the list of available ones. HD orders placed the day before, must be delivered by the end of the day. To reduce the workload of corporate drivers for the delivery process of HD orders, PL customers may be employed. At the time of order placement, PL customers are presented with a list of pre-generated HD order bundles, where they can bid on a subset. If a PL customer's bid wins, the associated bundles of HD orders are delivered to the parcel locker for the winning PL customer to collect along with their orders, which they then deliver to the respective HD customers in return for store credits. This problem is stochastic due to the dynamic nature of PL order arrivals. Hence, in each period, the retailer decides which bids to accept, which HD orders to dispatch, and which ones to postpone, aiming at minimizing the total operational costs. Further, routes are planned for vehicles, in each period, to deliver selected HD orders to customers and PL orders, along with those in the accepted HD bundles, to designated PLs before customer-chosen periods. To predict future PL orders and crowd-shipping capacity, we develop a stochastic look-ahead approach.

4 - Towards Equitable Workload Distribution in Last-Mile Delivery for the Gig Economy: The Dispatch Zone-Wave Problem

Abhay Sobhanan, University of South Florida, Tampa, FL, United States, Hadi Charkhgard, Iman Dayarian

Last-mile delivery (LMD) remains consistently crucial, constituting a significant expense within the logistics framework. In response to the escalating demand for faster delivery services, many retailers are turning to independent crowd couriers for order fulfillment. Crowdsourced drivers differ in terms of the time they become available, the capacity of their vehicles, and their next destinations, among other factors. Therefore, while crowdshipping revolutionizes LMD, its distinguishing factors introduce new operational challenges. One significant challenge is ensuring uninterrupted operations in crowdshipping, which is influenced by drivers' satisfaction in participating frequently in delivery tasks. Satisfied drivers are more likely to undertake subsequent jobs, maintaining operational flow. To ensure driver satisfaction, an equitable workload assignment is crucial since it directly impacts the compensation paid to and the cost incurred by each driver. In this talk, we propose a practical equity-aware framework tailored to LMD in crowdshipping environments. Using the proposed framework, several key research questions will be explored.

5 - Touting occasional drivers for mid-haul delivery

Margaretha Gansterer, University of Klagenfurt, Klagenfurt, Austria, Simona Mancini, Chefi Triki

Outlet villages offer online purchasing services combined with home delivery. Aiming at efficient last mile delivery, in our study we propose to exploit in-store customers to perform mid-haul deliveries on their way back home. The drawback of the system is that the presence of in-store customers willing to serve a specific area is strongly affected by uncertainty. In order to decrease the negative effect of this uncertainty, we propose to adopt a touting strategy to incentivize potential in-store customers to come to the outlet on a specific day and to perform, if needed, a mid-haul delivery, by offering them a discount voucher. In addition, the company may exploit customers who express their interest to act as occasional drivers, directly on site. Such customers may make bids to cover one or more areas. After receiving all the bids, the company decides through an auction system which ones to accept. The problem is formulated as a two-stage stochastic problem. We apply the model to instances based on a real-world case involving an outlet village located in the North-West of Italy. The obtained results show the benefit of applying the proposed mixed distribution system rather than making use of the owned fleet only.

TC37

Summit - 429

Nonlinear Learning and Optimization

Contributed Session

Chair: Yijin Ni, Georgia Institute of Technology, 10 Perimeter Summit Blvd NE, Brookhaven, GA, 30319, United States

1 - A Uniform Concentration Inequality for Mmd: Robustness to Functional Transformations

Yijin Ni, Georgia Institute of Technology, Brookhaven, GA, United States, Xiaoming Huo

Embedded in the *reproducing kernel Hilbert space* (RKHS), *maximum mean discrepancy* (MMD) is a metric for probability measures. Being successfully applied in a wide range of problems, the optimized MMD-based estimators through the choice of the kernel are proven to be consistent.

In this paper, we construct a uniform concentration inequality for the unbiased *U*-statistic estimate of the squared MMD, showing the uniform consistency of MMD from another perspective. More specifically, given a functional class *G* equipped with finite Gaussian complexity, we show that for any functional transformation g in class *G*, considering the transformed pair of probability metrics, the difference between the *U*-statistic estimate and the expected squared MMD has a probability bound that is irrelevant with the specific function *g*. In optimization problems that involve MMD in their objectives, through the choice of functional transformation applied to the involved random variables, the uniform convergence bound confirms the consistency of the MMD-based estimators in minimization, maximization, and minimax problems. Considering different choices of the functional class *G*, the application scenarios of the uniform concentration inequality include: (i) the bandwidth tuning phase of MMD in two-sample tests; (ii) the construction of a generative model through MMD in the case of model contamination; (iii) the choice of a Grassmannian in dimension reduction problems based on MMD; (iv) the independent component analysis problem based on MMD; (v) the machine learning models of MMD, such as generative adversarial networks, variational auto-encoders.

2 - Concurrent stochastic optimization and derivative estimation of quantiles for chance constrained problems

Felisa Vazquez-Abad, Hunter College-CUNY, Manhattan, NY, United States

Optimization under probability (or chance) constraints is known to be a challenging problem. Generalizing our preliminary results for one-dimensional problems, we study here the general model where the constraints may depend on a stochastic process parametrized by the control variable of the optimization problem. The distribution of the underlying process is not known and only samples of it are available, so the optimization algorithm must be built for streaming data. Instead of formulating the constraints as probability contraints, we use the alternative form in terms of quantile constraints, which is known to yield more efficient gradient-based algorithms. Common methods for solving such problems make use of advanced statistical functional estimation of the distribution function directly, using approximations that are provably convergent. Such approximations provide explicit functions of the control parameter that can be then used to build consistent estimators of derivatives of the quantile with respect to the control parameter. One of our main contributions is to propose a dynamic estimator of the derivative of the quantile directly (without using approximations of the distribution) that can be implemented as the optimization updates take place. Because quantiles are commonly estimated using the complete history of samples obtained, the computational challenges include an undesirable super-linear running time. We show that older samples can be re-used when updating the control parameters in the stochastic approximation algorithm, rather than having to re-start the estimation. We explicitly take account of the algorithmic complexity of the numerical methods for solving the stochastic optimization problem.

3 - Learning Regions of Interest for Bayesian Optimization with Unknown Constraints

Fengxue Zhang, University of Chicago, Chicago, IL, United States, Yuxin Chen

Optimizing objectives under constraints, where both the objectives and constraints are black box functions, is a common scenario in real-world applications such as the design of medical therapies, industrial process optimization, and hyperparameter optimization. One popular approach to handle these complex scenarios is Bayesian Optimization (BO). However, when it comes to the theoretical understanding of constrained Bayesian optimization (CBO), the existing framework often relies on heuristics, approximations, or relaxation of objectives and, therefore, lacks the same level of theoretical guarantees as in canonical BO. In this paper, we rely on the insight that optimizing the objective and learning the constraints can both help identify the high-confidence ***regions of interest*** (ROI) that potentially contain the global optimum. We propose an efficient CBO framework that intersects the ROIs identified from each aspect to determine the general ROI. Then, on the ROI, we optimize the acquisition functions, balancing the learning of the constraints and the optimization of the objective. We showcase the efficiency and robustness of our proposed CBO framework through the high probability regret bounds for the algorithm and extensive empirical evidence.

4 - Wasserstein-based kernels for unsupervised learning

Alfredo Oneto, ETH Zurich, Zurich, Switzerland, Blazhe Gjorgiev, Giovanni Sansavini

In numerous data clustering applications, the representation of objects often extends beyond vectorial data, posing a challenge. The bag-of-vectors representation has emerged as a versatile approach for encapsulating complex objects across various research domains. Additionally, leveraging tools based in computational optimal transport theory, particularly the Wasserstein distance, offers a robust geometric framework for treating these objects as probability distributions. This study explores the utilization of kernel methods harnessing the Wasserstein distance to craft a clustering framework tailored to such objects. Initially, we present a method for efficiently approximating Wasserstein distances via linear optimal transport. Subsequently, a kernel based on the Wasserstein distance is developed, ensuring its positive definiteness. Moreover, we integrate two clustering validity indices, which facilitate the comparison of clustering results obtained using kernels with varying parameters. To evaluate our framework, experiments are conducted on extensive datasets comprising power distribution graphs, employing kernels defined over both bags-of-vectors and vectorial data. The findings validate the efficacy of our proposed framework in adeptly handling complex objects within a clustering context.

TC38

Summit - 430

Drone Routing

Contributed Session

Aviation Applications

Chair: Ang Li, The Hong Kong Polytechnic University

1 - Mothership-based drone routing optimization via branch-and-price-and-cut

Reza Mirjalili, University of Houston, Houston, TX, United States, Gino Lim, Taewoo Lee

This paper presents a mothership-based drone routing problem to deliver parcels or visit waypoints for surveillance. Mothership (e.g., vehicle) follows a given sequence of fixed locations while operating drones onboard. Drones are allowed to serve multiple locations (i.e., customers) on a single flight. Small drones offer speed and cost advantages over ground vehicles. However, they face operational constraints such as limited payload and battery duration. Therefore, this paper aims to optimize the schedule of the mothership and drone flight paths by minimizing the total drone flight duration and waiting times of drones and the vehicle. Because solving the optimization model is computationally challenging, we propose a branch-and-price-and-cut (BPC) method, utilizing column generation to generate feasible flights and assignments. We introduce two valid inequalities to strengthen the lower bound of BPC: one employing dual optimal values for a stronger lower bound on flight numbers, and another mitigating the Big-M constraint effect for accurate waiting time calculation. Unlike existing BPC pruning policies based on LP relaxation, we propose a stronger bound based on a branching strategy, resulting in the reduction of the BPC tree size. Results demonstrate ~1.9-5% average gap improvement per node with the proposed bound and ~0.01-0.2% average gap improvement with the proposed cuts for 10-40 customer instances.

2 - Optimizing Vaccine Access: Equity-Focused Drone Delivery in Disease-Prone Region

Hamid Sayarshad, Cornell University, Ithaca, NY, United States

This research presents a novel framework that combines dynamic disease modeling and an optimization approach to enhance the fairness and equity of vaccine distribution through drone-based delivery. The objective is to ensure equal access to vaccines across different regions, including rural areas and small cities. Our approach seeks to achieve an optimal allocation of vaccines by considering regional infection rates and vaccination coverage, enabling effective prediction of vaccine demand. To accomplish this, we utilize a region-specific, dynamic disease model that factors in population size, infection rates, and vaccination rates. By integrating this dynamic disease model with a well-designed delivery network, we minimize travel and healthcare costs resulting from insufficient vaccination coverage, while also ensuring fair distribution. Our model takes into account logistical considerations specific to drone vaccine delivery, such as payload capacity, flight range, and regional vaccine demand. These considerations are crucial in addressing the unique challenges faced by rural areas and small cities in accessing healthcare services. To validate our model, we conducted a study on drone-based vaccine delivery during the COVID-19 pandemic, focusing specifically on Orange County and small cities. The findings of this study have significant practical implications for the development and implementation of drone-based vaccine delivery systems that prioritize fairness and equitable access to vaccines, particularly in smaller cities and rural areas.

3 - Drone-truck Arc Routing Problems

Sung Hoon Chung, Binghamton University, Binghamton, NY, United States, Emad Alenany

We address a Drone-Truck Arc Routing Problem (DT-ARP) for which an integer programming model is developed based on the modeling concept called an operation. The considered problem involves adding a new drone feature to the rural postman problem. A heuristic approach is proposed, where local search and simulated annealing methods are used to select initial truck tours. We also address the best combination of drone launch and recovery nodes to minimize the makespan. Numerical examples are presented to discuss the effectiveness of the proposed model and heuristic for DT-ARP.

4 - A Two-Echelon Bidirectional Energy Supply Problem with Uncrewed Electric Aerial and Ground Vehicles

Hyunhwa Kim, University of Illinois Urbana-Champaign, Urbana, IL, United States, Eleftheria Kontou

In emergency response settings or contested environments, energy supply to various demand nodes is constrained due to damaged or insufficient power infrastructure. Electric uncrewed ground (UGVs) and aerial vehicles (UAVs) can play a notable role in providing backup power, discharging energy directly to meet demand and recharging en-route if necessary. This research aims to design a two-echelon multimodal logistics system for bidirectional energy supply. The first echelon is distributing deployed UGVs and UAVs from a depot to satellite locations. In the second echelon, these vehicles travel from satellites to meet energy demand with discharging their batteries and recharge en-route if needed. We formulate this system as a bidirectional electric vehicle location-routing problem, which we solve using column generation to optimize vehicle routes and determine satellite locations to minimize energy consumption of the system. We conduct numerical experiments on transportation networks and perform sensitivity analysis, assessing the impact of the battery capacities of UGVs and UAVs on the optimal energy supply.

5 - Congestion Sensitive Strategies for Drone-based Multimodal Last-mile Urban Parcel Delivery

Ang Li, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Urban parcel delivery has emerged as a high growth market, and the resulting delivery traffic can pose great challenges in dense urban areas. There is growing interest in supplanting the conventional model of a dedicated delivery person operating a van to alternatives featuring new classes of vehicles such as drones, autonomous ground vehicles, cargo bikes and non-motorized vehicles. This work proposes combined delivery strategies using trucks, cargo bikes and drones. We develop and compare multi-modal delivery strategies with various scenarios in both single and multiple echelon last-mile delivery network. We plan the delivery route, schedule, mode selection, facility location, task assignment in the developed optimization models. Then, a congestion model is developed and integrated to the multimodal delivery model through Macroscopic Fundamental Diagram (MFD) and traffic simulations. The congestion model relates multimodal delivery strategies and congestion effects on road network. We evaluate the benefit of multi-modal delivery in both uncongested and congested transportation networks. Results show that delivery models with multiple vehicles modes in both single- and multi- echelon networks are more efficient in terms of total delivery cost than traditional truck only scenario. The multi- modal delivery strategies in two- echelon networks outperform other strategies in extremely congested situations. We suggest taking advantage of synergistic operation among emerging vehicle types, especially drones for more efficient parcel delivery.

TC39

Summit - 431

Optimizing Service Networks and Systems

Contributed Session

Chair: arvind rathore, 14068

1 - An iterative commodity aggregation and disaggregation method for service network design problems

Louis Bonnet, LAAS-CNRS, Université de Toulouse, Toulouse INP, Toulouse, France, Simon Belieres, Mike Hewitt, Sandra Ulrich Ngueveu

Service Network Design Problems (SNDP) have been widely studied in the OR literature but become rapidly intractable as the number of commodities increases. We aim to alleviate the impact that the number of commodities may have on the problem difficulty through the use of aggregation and disaggregation techniques. In this context, we define a relaxation of the original problem that is based on the idea of aggregating commodities, reducing the number of flow variables and, thus, the size of the problem. However, this reduction in size may lead to solutions that are not feasible to the original problem. To circumvent this issue, we propose an invalidity detection procedure that identifies the underlying reasons for this infeasibility. We also introduce a disaggregation scheme that intends to tighten the solution space of the aggregated model by refining the information available. To control the size of the aggregated model, our scheme aims to solely disaggregate information on a subset of arcs for which it appears necessary. Ultimately, starting from the initial relaxation, we propose an iterative solution approach that (i) solves the aggregated SNDP, (ii) checks the validity of the solution, and (iii) disaggregates specific commodities over carefully chosen arcs until a feasible and thus, optimal, solution to the original problem is found.

2 - Welfare-Maximizing Designs of Two-Tier Multi-Server Service Systems

Jing Yang Xi, University of California, Berkeley, Berkeley, CA, United States, Wai Kin (Victor) Chan

Heterogeneous servers are often adopted to accommodate customers with different delay sensitivities. A widely-used model is the two-tier *express system* where the first tier provides free regular service with slower servers and the second tier provides paid express service with faster servers, i.e., impatient customers can pay extra to reduce delay. We analyze the social welfare and welfare-maximizing price of such systems and propose a *reward system* where the second tier now provides rewards (charges negative toll) to attract patient customers; in terms of welfare, this is equivalent to charging higher prices for slower servers. We provide the conditions for when the reward system generates more welfare than the express system and vice versa. This duality—existence of two systems—requires the social planner to solve for the maximum welfare of both systems to maximize total market welfare. To complement our analytical analysis, we collect data on passengers to model the performance of the express and reward system in airport security check. Paradoxically, the free regular service in the reward system may have a lower waiting time than the paid express service in the express system, leading to higher utility for every passenger and higher welfare in the reward system.

3 - Heuristics using Consolidation Patterns for Increasing Solution Speeds of Service Network Design Problems using the Rolling Horizon Solution Method

Ryan Tramp, University of Alabama, Tuscaloosa, AL, United States, Nickolas Freeman

The service network design problem (SNDP) is an NP-hard networking problem that has proven to be computationally challenging in its most basic form. Thus, it is difficult to explore additional problem aspects, such as uncertainty in demand and order arrival, for all but very small instances. This research explores order consolidation patterns that emerge in optimal solutions to small SNDP instances and uses these insights to generate heuristics that can significantly reduce solution times while maintaining low gaps from optimality. An analysis of the optimal solution for small instances reveals that orders tend to consolidate around one of the locations where orders arrive at the network or order destinations depending on several factors. A major factor is the number of orders reaching a critical threshold that makes movement worthwhile. We encode common characteristics of solutions when the number of orders is around this threshold into our heuristics. Preliminary testing shows parity in solution qualities when comparing the heuristics to a rolling horizon approach that utilizes MIP models, with the heuristics offering significant reductions in computation time.

4 - Limiting Route Predictability using a Linear Approximation of Shannon Entropy

arvind rathore, University at Buffalo, Amherst, NY, United States, Robert Dell, Rajan Batta

This paper tackles route predictability by integrating unpredictability as a core driver of route design. We create a Multiple Patroller Math-heuristic based on traditional Mixed Integer Linear Programming (MILP) models that include a linear approximation of change in entropy as a constraint or an objective for improving entropy over subsequently generated route. Prior publications only utilize entropy for route selection. Applying our math-heuristic to a real-world scenario from the southeastern police district of Baltimore City, we discover a significant increase in entropy and coverage while maintaining a practical balance with a prize-collecting objective. We also develop cuts to improve solution time. The case study not only illustrates the viability of our model but also brings into focus its potential for widespread application in various secure routing contexts. We extend this work by demonstrating arrival time diversification, a common objective in Cash-in-Transit problems, and observe unpredictably scattered visit times with and without time windows.

TC40

Summit - 432

Feedback and Information in Human-Algorithm Interactions

Invited Session

Behavioral Operations Management

Chair: Stephen Leider, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Clare Snyder, Michigan Ross, Ann Arbor, MI, United States

1 - Algorithm Reliance for Binary Classification: Implications for Accuracy and Fairness

Clare Snyder, Michigan Ross, Ann Arbor, MI, United States, Samantha Keppler, Stephen Leider

In this research we use laboratory experiments to answer the question: how effective will binary classification algorithms designed for fairness (i.e., equal opportunity) be when they are channeled through a human decision-maker? The design of fair AI has received considerable attention in recent years, but another stream of research shows that people often under-rely on algorithms, even when these algorithms are proven to be highly accurate. It is therefore possible that when humans have ultimate decision authority, they will deviate from or ignore algorithms' advice in ways that undermine equal opportunity or other fairness outcomes. This may be particularly true for settings in which feedback is asymmetrically censored (e.g., loans, admissions, diagnoses) – where verification bias is present. In these settings, algorithms that achieve fairness by relaxing the threshold for one population may appear very different from algorithms that achieve fairness by doing the opposite, *even when the two algorithms are equally accurate.* Our results have implications for fair algorithm design.

2 - Warnings and Endorsements: Improving Human-AI Collaboration in the Presence of Outliers

Matthew DosSantos DiSorbo, Harvard Business School, Boston, MA, United States, Kris Ferreira, Maya Balakrishnan, Jordan Tong

1. Problem definition: While artificial intelligence (AI) algorithms may perform well on data that are representative of the training set (inliers), they may err when extrapolating on non-representative data (outliers). How can humans and algorithms work together to make better decisions when faced with outliers and inliers?

2. Methodology/results: We study a human-AI collaboration on prediction tasks using an anchor-and-adjust framework, and hypothesize that humans are biased towards naive adjustment behavior: making adjustments to algorithmic predictions that are too similar across inliers and outliers, when ideally adjustments should be larger on outliers. In an online lab experiment, we demonstrate that participants are indeed unable to sufficiently differentiate absolute adjustments to an AI algorithm when faced with both inliers and outliers, leading to a 143-176% increase in absolute deviation from the optimal prediction compared to participants who only face either all inliers or all outliers. We design a 'warning' ('endorsement') that alerts participants when feature values constitute outliers (inliers) and, in another experiment, we show that this warning (endorsement) helps participants differentiate adjustments, ultimately reducing absolute deviation from the optimal prediction by an

average of 35% (28%). However, we demonstrate that deploying both interventions together reduces participants' absolute deviation from the optimal prediction by 49%.

3. Managerial implications: Our work uncovers a behavioral bias towards naive adjustment behavior, and identifies a simple, educational intervention that mitigates this bias. Ultimately, we hope that this work will help managers best equip their employees with the knowledge they need to succeed in human-AI collaborations.

3 - Enhancing the Wisdom of Crowds when Experts use Algorithmic Advice

Angshuman Pal, Indiana University Kelley School of Business, Bloomington, IN, United States, Asa Palley, Ville Satopaa

In many managerial settings, good decision making depends on obtaining an accurate forecast of an uncertain variable of interest. Useful knowledge about the variable may be accessible to both human experts and AI models, where each may have their own relative advantages in forming an accurate forecast. To make use of both sources of information, AI advice can be provided to human experts who can assess how much weight it should be given when updating their prior forecasts.

Since useful information may be held by different individuals, combining forecasts from multiple experts can also boost the accuracy of the aggregate forecast. However, while providing AI advice to humans can be helpful at an individual level, it may also induce correlations in judgment errors that hamper efficient combination of information at the aggregate level.

Using a stylized Bayesian model of information held by human experts and the algorithm, we derive a new procedure for aggregating judgments when humans receive AI advice. The method uses individual responses to estimate how much weight to put on human versus AI forecasts, and forms an aggregate forecast by taking the final average human forecast and adjusting it toward or away from both the initial average human forecasts and the average AI advice.

4 - Productivity Implications of Working Side Jobs

Na Hyun Kim, University of Pennsylvania, Philadelphia, PA, United States, Park Sinchaisri

With the success of online platforms that utilize flexible forms of labor, it has never been easier to hold multiple jobs. However, research into such workers have been sparse namely due to difficulty in identification and heterogeneity of jobs to allow for an adequate comparison. We study a particular class of platform workers, medical professionals that work on third-party telehealth platforms, and mimic their work setting in an online game to measure productivity spillovers between the worker's main job and side job on said platform. We confirm our findings with surveys and interviews of such workers on a major telehealth platform in the US.

TC41

Summit - 433

Advances in Queueing Theory

Invited Session

Applied Probability Society

Chair: Martin Zubeldia, University of Minnesota, Minneapolis, MN, United States

1 - Electrifying the Fleet: Non-Asymptotic Dimensioning of EV-Based Mobility Systems with Performance Guarantees

Saif Benjaafar, University of Michigan, Ann Arbor, MI, United States, Jingyuan Wan, Martin Zubeldia

We consider the problem of dimensioning an on-demand mobility service that relies on a fleet of electric vehicles and a dedicated charging infrastructure. We model the system as a closed queueing network and derive bounds on the minimal number of vehicles that guarantee a target availability and a charging capacity that guarantees that expected delay at the chargers does not exceed a target threshold. The bounds, which hold for any sized system, yield tight non-asymptotic dimensioning rules.

2 - Exponential Tail Bounds on Queues: a Confluence of Non-Asymptotic Heavy Traffic and Large Deviations

Siva Theja Maguluri, ISyE Georgia Tech, Atlanta, GA, United States, Prakirt Jhunjhunwala, Daniela Hurtado

Obtaining the exact steady-state distribution of queue lengths is not feasible in general. Therefore, we establish bounds for the tail probabilities of queue lengths. Specifically, we examine queueing systems under Heavy-Traffic (HT) conditions and provide exponentially decaying bounds for the probability $P(\epsilon q>x)$, where ϵ is the HT parameter denoting how far the load is from the maximum allowed load. Our bounds are not limited to asymptotic cases and are applicable even for finite values of ϵ, and they get sharper as $\epsilon\to 0$. Consequently, we derive non-asymptotic convergence rates for the tail probabilities. Unlike other approaches such as moment bounds based on drift arguments and bounds on Wasserstein distance using Stein's method, our method yields sharper tail bounds. Furthermore, our results offer bounds on the exponential rate of decay of the tail, which can be interpreted as non-asymptotic versions of Large Deviation (LD) results.

We demonstrate our approach by presenting tail bounds for a continuous time Join-the-shortest queue (JSQ) load balancing system. We not only bridge the gap between classical-HT and LD regimes but also explore the large system HT regimes, where the system size and the system load increase simultaneously. Our results also close a gap in the existing literature on the limiting distribution of JSQ in the super-NDS (a.k.a. super slowdown) regime. This contribution is of an independent interest. Here, a key ingredient is a more refined characterization of state space collapse for JSQ system, achieved by using an exponential Lyapunov function designed to approximate the sup norm.

3 - Matching Impatient and Heterogeneous Demand and Supply While Learning

Amy Ward, The University of Chicago Booth School of Business, Chicago, IL, United States, Weiliang Liu, Zhang Xun

We study a two-sided network where heterogeneous demand (customers) and heterogeneous supply (workers) arrive randomly over time to get matched. Customers and workers arrive with a randomly sampled patience time (also known as reneging time in the literature), and are lost if forced to wait longer than that time to be matched. The system dynamics depend on the matching policy, which determines when to match a particular customer class with a particular worker class. The issue in developing a matching policy is that model primitives are unknown. Our objective is to develop a policy that has small regret, where regret is the difference between the cumulative value of matches, minus costs incurred when demand and supply wait, between a proposed policy, that does not have knowledge of model primitives, and a benchmark policy, that does have full knowledge of model primitives. Our benchmark policy is an asympotically optimal policy (on fluid scale, as demand and supply rates grow large). A key challenge is that the benchmark policy depends on the patience time distributions, and may be different for different distributions, even when the mean is the same.

4 - Heavy-Traffic Optimality of Skip-the-Longest-Queue in Parallel Service Systems

Martin Zubeldia, University of Minnesota, Minneapolis, MN, United States

We consider a parallel service system consisting of a fixed number of single server queues, where a single stream of jobs arrive to it and need to be dispatched immediately upon arrival to one of the queues. For this setting, we propose a new dispatching policy dubbed Skip-the-Longest-Queue (SQL), where incoming jobs are sent to a queue that is not the longest one. We show that this policy is asymptotically optimal in heavy-traffic, even when the ID of the longest queue is only updated once every many arrivals.

TC42

Summit - 434

Probability on Networks

Invited Session

Applied Probability Society

Chair: Souvik Dhara, Purdue University, West Lafayette, IN, United States

1 - Convergence of Metropolis chains on large stochastic block models

SOUMIK PAL, University of Washington, Seattle, WA, United States, Siva Athreya, Raghav Somani, Raghav Tripathi

Metropolis Markov chains are commonly used to sample from dense random graph models. However, almost nothing is known about their rate of convergence to equilibrium. We consider the case when the size of the graph goes to infinity. In this regime, paths of a slightly modified Metropolis chain converges to a deterministic curve on the space of graphons. This curve is close to the gradient flow of the Hamiltonian on the metric space of graphons. A geometric analysis of this curve along this line leads to an exponential rate of convergence under a notion of convexity of the Hamiltonian.

2 - A Graph Limit Perspective on Sampling-Based GNNs

Yeganeh Alimohammadi, Stanford University, Stanford, CA, United States, Luana Ruiz, Amin Saberi

With the increasing size and complexity of graph data, scaling graph neural networks (GNNs) has become a crucial challenge. Over the past few years, many GNN training algorithms based on sampling have emerged to address this issue. In this talk, I will present a unified theoretical framework for training Graph Neural Networks on large input graphs using small, fixed-size sampled subgraphs. This framework is applicable to a wide range of models, including popular sampling-based GNNs such as GraphSAGE and FastGCN. Leveraging the theory of graph local limits, we demonstrate that, under mild assumptions, parameters learned from training sampling-based GNNs on small samples of a large input graph fall within an ϵ-neighborhood of the outcomes achieved by training the same architecture on the entire graph. We derive bounds on the number of samples, the size of the graph, and the training steps required as a function of ϵ. Our results provide a theoretical understanding of using sampling in training GNNs and suggest that by training GNNs on small samples of the input graph, practitioners can more efficiently identify and select the best models, hyperparameters, and sampling algorithms.

3 - Combining Rollout Designs and Clustering for Causal Inference with Network Interference

Mayleen Cortez, Cornell University, Ithaca, NY, United States, Matthew Eichhorn, Christina Yu

Estimating causal effects under *network interference* is pertinent to many real-world settings. However, the underlying interference network is often unknown to the practitioner, which precludes many existing techniques that leverage this fine-grained information. A recent line of work uses staggered rollout experimental designs, polynomial interpolation and low-order interaction models to obtain unbiased estimators *that require no network information*. However, the required polynomial extrapolation can lead to prohibitively high variance when the polynomial degree is higher than 1. In this work, we propose a two-stage staggered rollout experimental design that can use, but does not require, coarse-grained network knowledge to sub-sample the population and roll out treatment to a greater proportion of this sub-population. This greatly reduces variance from extrapolation, but at the cost of incurring some bias and sampling variance. Through numerical simulations, we explore the trade-off between the error attributable to the subsampling of our experimental design and the extrapolation of the estimator for both synthetic and real-world networks. We also present a general and explicit expression of the bias of the estimator under the proposed design as well as a bound on the variance. We find that when the degree is greater than 1, the proposed design greatly reduces error and outperforms against baseline estimators, especially when the treatment probability is small.

4 - Kernel and Graphical Methods for Conditional Inference

Bhaswar Bhattacharya, University of Pennsylvania, Philadelphia, PA, United States, Anirban Chatterjee, Ziang Niu

In this talk we will discuss methods for conditional nonparametric inference, such as conditional 2-sample and goodness-of-fit testing, using kernels and nearest-neighbor graphs. Leveraging ideas from nonparametric statistics and random geometric graphs, we will construct tests that can be readily applied to a range of problems, such as model calibration, regression curve evaluation, and validation of emulators in simulation-based inference.

TC43

Summit - 435

Chang-Han Rhee session

Invited Session

Applied Probability Society

Chair: Chang-Han Rhee, Northwestern University, Evanston, IL, United States

1 - Higher-criticism for multi-sensor sparse change-point detection

Yao Xie, Georgia Institute of Technology, Atlanta, GA, United States, Tingnan Gong, Alon Kipnis

We present a procedure based on higher criticism (Dohono \& Jin 2004) to address the sparse multi-stream quickest change-point detection problem. Namely, we aim to detect a change in the distribution of the multi-stream that might occur in some unknown time instant. The change itself, if occurs, only affecting a few data streams out of potentially many while those affected streams, if exist, are unknown to us in advance. Our procedure involves testing for a change point in individual streams and combining multiple tests using higher criticism. As a by-product, our procedure also indicates a set of streams suspected to be affected by the change. We characterize the information-theoretic detection limit under a Gaussian heteroscedastic mean shift when individual tests are based on the LR or GLR and show that our procedure attains this limit; this limit coincides with existing results in the homoscedastic case. We demonstrate the effectiveness of our method compared to other procedures using numerical evaluations.

2 - DPZero: Private Fine-Tuning of Language Models without Backpropagation

Sewoong Oh, University of Washington, Seattle, WA, United States

The widespread practice of fine-tuning large language models (LLMs) on domain-specific data faces two major challenges in memory and privacy. First, as the size of LLMs continues to grow, the memory demands of gradient-based training methods via backpropagation become prohibitively high. Second, given the tendency of LLMs to memorize training data, it is important to protect potentially sensitive information in the fine-tuning data from being regurgitated. Zeroth-order methods, which rely solely on forward passes, substantially reduce memory consumption during training. However, directly combining them with standard differentially private gradient descent suffers more as model size grows. To bridge this gap, we introduce DPZero, a novel private zeroth-order algorithm with nearly dimension-independent rates. The memory efficiency of DPZero is demonstrated in privately fine-tuning RoBERTa and OPT on several downstream tasks.

3 - Diffusion Models for Generative Optimization via High-Dimensional Modeling

Minshuo Chen, Northwestern University, Evanston, IL, United States

Deep generative AI, e.g., diffusion models, achieves state-of-the-art performance in various high-dimensional data modeling tasks. Such empirical successes have been challenging conventional wisdom. In this talk, we will focus on diffusion models to explore their methodology and theory. We will first understand how diffusion models efficiently model complex high-dimensional data, especially when there are low-dimensional structures in them. Then, we leverage our understanding of diffusion models to motivate a next-generation optimization method, termed “generative optimization”. Specifically, we utilize diffusion models as a data-driven solution generator to an unknown objective function. We propose a learning-labeling-generating algorithm incorporating the targeted function value as guidance to the diffusion model. Theoretically, we show that in the offline setting, the generated solutions yield large function values on average. Meanwhile, the generated solutions closely respect the data intrinsic structures in the training set. Empirically, we demonstrate a good synergy of generative optimization with reinforcement learning.

4 - Accelerating diffusion models with stochastic Runge-Kutta method

Yuchen Wu, University of Pennsylvania, Philadelphia, PA, United States

In this talk, I will present a new framework for accelerating diffusion models using the stochastic Runge-Kutta method. Our approach is different from the previous acceleration methods, and offers an improved theoretical guarantee. Numerical experiments confirm the practicality of our approach.

TC44

Summit - 436

Emergency Services Efficiency and Workload Balance

Contributed Session

Chair: Jenna Ercolani, Wichita State University, 1845 Fairmount St, Wichita, KS, 67260, United States

1 - A Dynamic Programming Approach to Balancing Workload in EMS Through Intra-Shift Crew Swapping

Setareh Darvishi, Wichita State University, Wichita, KS, United States, Laila Cure

This project develops an optimization model for reassigning ambulance crews to different stations at different points in time throughout a shift to balance their workload. The collaborating Emergency Medical Services (EMS) system currently assigns crews to stations for a full 12-hour shift. These fixed assignments often result in an imbalanced workload among crews given that demands at different locations can be significantly different, even during the same shift. We propose a stochastic dynamic programming formulation in which the state of the system at each decision epoch is described by the current station assignment and cumulative workload of all crews in the system. The proposed state information can be extracted in real time from dispatch data commonly available to EMS systems and the main source of uncertainty is the random nature of 9-1-1 calls. The optimal policy aims at minimizing the expected total cost. The terminal cost function measures the disparity in the expected terminal workload of all crews and the intermediate cost function represents the undesirable aspects of swapping. The applicability of the proposed strategy and the effectiveness of the model will be evaluated using data from Sedgwick County EMS.

2 - Spatiotemporal Analysis of Volunteer First Responders Effectiveness

Michael Khalemsky, Hadassah Academic College, Jerusalem, Israel, Janna Ataiants, Stephen Lankenau, Anna Khalemsky, Alexis Roth, Gabriela Marcu, David Schwartz

Rapid first aid is vital for reducing mortality and improving long-term prognosis during medical emergencies. One approach to achieving faster response times is through smartphone-based Volunteer First Responder (VFR) networks. VFR effectiveness is measured by factors such as arrival time, relevant intervention, and medical outcomes. Geographic coverage is a commonly used measure of emergency medical service effectiveness. This study presents a novel analytical technique for analyzing VFR network effectiveness in terms of geographic coverage using real-world data from a field study. The analysis considers the unpredictable location and availability of VFRs during an emergency event, aiding decision-making for recruitment and retention efforts and improving collaboration with EMS. Findings can enhance VFR network administrators' efforts to save lives.

3 - Optimizing Police Staffing Allocation Using Data Envelopment Analysis

Andrew Hall, Marymount University, Arlington, VA, United States, Erik Alda

Efficiently allocating police personnel to match workload demands is a critical challenge facing law enforcement agencies nationwide (Wilson & Weiss, 2014). Understaffed departments struggle with slow response times, officer safety risks, and reduced community engagement, while overstaffing wastes limited resources (Wilson & Grammich, 2012; Chalfin & McCrary, 2018). The uneven distribution of crime and calls for service within jurisdictions further complicates staffing decisions (Wu & Murray, 2005; Wheeler, 2022). Traditional staffing models based on population ratios or historical precedent often fail to capture the dynamic, place-specific nature of police workload (Wilson & Weiss, 2012). This can result in the overpolicing of some communities and the underprotection of others (Kennedy, 1997; Mastrofski, 2004). Inequitable deployment hinders police effectiveness and undermines police legitimacy and community trust (Rosenbaum & Schuck, 2012). The purpose of this study is to optimize the deployment of a small city police patrol resources in using frontier methods and its variants. Working on a research-practitioner partnership between the police department and Marymount University, we develop, implement, and evaluate a data-driven staffing allocation model that enhances the efficiency and equity of police services. We will develop a data envelopment analysis (DEA) model by integrating data on labor, capital, and outputs produced. Various efficiency models, including traditional DEA and advanced variations like inverse data envelopment analysis (efficiency-based model), Malmquist-Luemberger Index, and partial frontier models, will be used to assess efficiency. We also implement changes to police allocation based on the efficiency model results.

4 - Physics-Based Surrogate Model Development for Prediction of Crew Member Workload in Emergency Medical Services

Jenna Ercolani, Wichita State University, Wichita, KS, United States

Workload in Emergency Medical Services (EMS) has been historically measured by metrics such as call volume or Unit Hour Utilization (UHU). This paper investigates a physics-based surrogate model to calculate a measure of an EMS crew member’s utilization more accurately, considering indirect work tasks, such as documentation time and shift start activities. The true measure of utilization was based on a trace-based simulation output incorporating VACP (Visual, Auditory, Cognitive, Psychomotor) workload measures built with real-world data and estimates for indirect work. The constructed surrogate models, and conventional measures of workload were tested against the simulation outputs for a years’ worth of data. Root mean square error (RSME), symmetric mean absolute percent error (sMAPE), and Pearson correlation estimates were used to compare models. The model with the optimal metrics included shift start time and a call duration estimate including averages of driving to post time, documentation time, and call response time. The optimal model was tested using sensitivity analysis and uncertainty calculation. General models were given for use in other EMS systems.

TC45

Summit - 437

Empirical Healthcare Operations

Invited Session

Health Applications Society

Chair: Zahra Mobini, Georgia Tech, Atlanta, GA, United States

Co-Chair: Wilson Lin, Santa Clara University, Santa Clara, CA

1 - Promoting Generics: Effects on Pharmaceutical Quality

In Joon Noh, Korea University, Seoul, Korea, Republic of, Hessam Bavafa, Christian Blanco

Generic drugs are a cornerstone of affordable healthcare. We examine the pharmaceutical quality effects of the Generic Drug User Fee Amendments (GDUFA), a hallmark legislation enacted by Congress in 2012 that armed the FDA with resources to improve the timeliness of generic drug application reviews. GDUFA succeeded in shortening the total application review time, which includes waiting time, from a median of 44 to 14 months. We find, however, that generic drugs approved after GDUFA have 28.6% higher recall rates. We show that this effect has been persistent since the second year of the law’s implementation, and that the increase in recall rates was driven by manufacturing-related deficiencies, specifically, contamination issues. We also investigate heterogeneity by recall severity and find that the increase in recalls after GDUFA is driven by severe (Class I and II) recalls. These findings inform previously unquantified quality reductions stemming from the policy promoting generic drugs.

2 - The Role of Physician Integration in Alternative Payment Models: The Case of The Comprehensive Joint Replacement Program

Kraig Delana, IE Business School, Madrid, Spain, Christopher Chen

U.S. physicians are increasingly unifying (both horizontally and vertically) while transitioning from fee-for-service to alternative payment models. We empirically examine the role of both horizontal and vertical integration of orthopaedic surgeons in driving heterogeneity in the impact of the Comprehensive Joint Replacement (CJR) alternative payment program. We find CJR hospitals with high horizontal and vertical integration see an increase in both hospital costs and risk-adjusted complication rates of 3.17% & 1.18, respectively. Moreover, we present evidence showing integration affects physician care decisions regarding length-of-stay and discharge disposition which explains the changes in hospital payments. Our findings provide novel insights into provider integration and clinical decisions that are relevant for policymakers, payors, and healthcare providers.

3 - Budget Blues in Healthcare: Unveiling the True Impact of Maryland's Hospital Payment Reform

Zahra Mobini, Georgia Tech, Atlanta, GA, United States, John McCandlish, Turgay Ayer

To rein in healthcare costs, Maryland implemented a global budget program (GBP) for its hospitals, establishing yearly cost growth targets for all inpatient and outpatient services. Leveraging an extensive claims dataset for commercially insured individuals, our study empirically examines how the GBP influences care delivery patterns within hospitals.

4 - Valuing Nursing Productivity in Emergency Departments

Sokol Tushe, University of South Florida, Tampa, FL, United States, Hao Ding, Diwas KC, Donald Lee

The United States is over a decade into the worst nursing shortage crisis in history, fueled by chronic underinvestment. To demonstrate to hospital managers and policymakers the benefits of investing in nursing, we clarify the positive downstream effects of doing so in the ED setting. We use a high-resolution data set of patient visits to the ED of a major U.S. academic hospital. Time-dependent hazard estimation methods (nonparametric and parametric) are used to study how the realtime service speed of a patient varies with the state of the ED, including the time-varying workloads of the assigned nurse. A counterfactual simulation is used to estimate the gains from increasing nurse staff in the ED. We find that lightening a nurse’s workload by one patient is associated with a 14% service speedup for every patient under the nurse’s care. Simulation studies suggest that adding one more nurse to the busiest 12-hour shift of each day can shorten stays and avert $160,000 in lost patient wages per 10,000 visits. The reduction in service times also frees up capacity for treating more patients and generates $470,000 in additional net revenues for the hospital per 10,000 visits. Extensive sensitivity analyses suggest that our key message—investing in nursing will more than pay for itself—is likely to hold across a wide range of EDs.

TC46

Summit - 438

Simulation Modeling in Healthcare

Invited Session

Health Applications Society

Chair: Sait Tunc, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Hyunwoo Shin, Virginia Tech, Blacksburg, VA, United States

1 - Enhancing Cost-Effectiveness of Screening for Chronic Obstructive Pulmonary Disease (COPD) with Smoking Cessation Interventions in China: A Simulation Modeling Study

Qiushi Chen, Penn State University, University Park, PA, United States, Yiwen Fan, Simiao Chen

China has the highest mortality rate from chronic obstructive pulmonary disease (COPD) in the world. Previous studies have shown the cost-effectiveness of COPD screening for the general population in China. Considering that smoking is the most common risk factor for COPD burden in the country, in this study, we aim to explore the impact of adding smoking cessation interventions following the screening to further enhance the management of COPD. We develop a microsimulation model and calibrate it to the estimates of the disease epidemiology and smoking behaviors in China. Using the calibrated simulation model, we evaluate multiple follow-up smoking cessation interventions by different intensities, including minimal counseling (counseling <90 min), intensive counseling (counseling >=90 min), and pharmacotherapy, and compare with screening only without cessation interventions. We find that all smoking cessation interventions are highly cost-effective and can further reduce exacerbations and deaths from COPD.

2 - Improving Vascular Access Outcome in Hemodialysis Considering Frailty: A Microsimulation Approach

Yiwen Cao, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen, Eugene Lin, Karen Woo, Mariana Murea

In hemodialysis, the choice of vascular access (catheter, arteriovenous [AV] fistula, or AV graft) influences maturation duration, length of functional usage, complication rates, and expected lifespan. While clinical guidelines generally recommend using AV fistulas, variation in patient physiologic function -- specifically frailty levels -- necessitates a more personalized recommendation for vascular access. With a team of clinicians, we developed a microsimulation model to evaluate the short-term and lifetime effects of different vascular access types on hemodialysis outcomes. Due to limited data directly relating to frailty, we employed various methods to estimate its impact on patient outcomes. We derived relative risks associated with access maturation from national prevalence data and existing literature and applied a risk multiplier to access failure using age as an indirect indicator, thereby simulating the impact of frailty. We observed the performance differences across vascular access choices and captured expected health outcomes with respect to patient characteristics including frailty and age. As multiple failures of vascular access are common in hemodialysis, we allowed for repeated decisions of vascular access choice when previous access failed in the model. We therefore stratified the model inputs by the number of prior failures and performed scenario analyses to provide cohort-specific recommendations for vascular access choice.

3 - Calibration of Simulated Organ Allocation Models Using Conformal Prediction Concepts

Hyunwoo Shin, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Xi Chen

Simulation models have played a crucial role in evaluating the effectiveness of organ allocation policies. However, traditional validation methods such as hypothesis testing and confidence intervals often fall short, primarily due to the limited and non-stationary nature of the transplantation data, compounded by the variability of p-values across simulation replications. Notably, p-values can exhibit significant variation even within the same model, depending on the number of outputs, thereby highlighting inconsistencies in traditional statistical testing. To address these challenges, our study introduces a novel approach inspired by the principles of conformal prediction. This method estimates the relative location of original data when presented with simulated data, offering a robust alternative to conventional tests. By adapting conformal prediction concepts, we provide a solution that enhances the calibration of simulation models and increases the reliability of their results, thereby supporting better-informed decision-making in organ allocation policies.

4 - Managing Cargo Drones for Health Product Delivery

Boas Meijer, INSEAD, Fontainebleau, France, Stephen Chick, Prashant Yadav

Drones are increasingly used for the last-mile delivery of health products. The value of drones as an alternative delivery method lies in their ability to respond quickly to medical emergencies. They also provide an efficient alternative to ground transport for less urgent replenishments of certain products, and in certain geographies. In this research, we study a drone service provider's decision on how to set its drone capacity, and schedule its drones in light of these two use cases. More specifically, we assume the decision maker takes into account differences between incoming orders for drone transport in terms of responsiveness needs, and delivery time. We establish and evaluate control policies to get a better sense of the tradeoff between responsiveness and capacity in this setting, and different levels of resource sharing can mitigate this tradeoff. Our models are calibrated using data on existing drone operations from our industry partner.

TC47

Summit - 439

Precision in Practice: Data Insights in Healthcare Management

Invited Session

Health Applications Society

Chair: Jacob Jameson, Harvard University, Cambridge, MA, United States

1 - Variation in Batch Ordering of Imaging Tests in the Emergency Department and the Impact on Care Delivery

Jacob Jameson, Harvard University, Cambridge, MA, United States, Soroush Saghafian, Robert Huckman, Nicole Hodgson

We use practice variation across physicians to uncover the role of test-ordering on care delivery in the ED. Using records of over 45,000 Mayo Clinic emergency department visits, we find that quasi-random assignment to a top (versus bottom) decile batching physician significantly increases excess diagnostic testing. Instrumental variable results show that batching leads to an additional 13.5 tests per 100 batches, resulting in excess tests that would not have been ordered under a sequential-ordering strategy. Contrary to expectations, batch ordering does not reduce ED length of stay or affect 72-hour return rates, challenging the perceived efficiency of batching and highlighting the necessity for more targeted testing strategies in emergency care.

2 - Mathematical models for cancer decision-making considering healthcare system and access constraints

Melissa Franco, Stanford, Stanford, CA, United States

Decision-makers in resourced-constrained settings face a unique barrier in implementing cancer control policies that reduce the time from symptom presentation to diagnosis and treatment to improve cancer outcomes. Mathematical modeling has increasingly been used to inform cancer control policy decisions; however, most modeling analyses are commonly tailored for more developed healthcare systems. These analyses overlook capacity constraints, which may bias the achievable impact of cancer health interventions and policies. The central component of this project will be to critically illustrate how ignoring or changing capacity constraints in mathematical models will impact cancer outcomes and, ultimately, healthcare decision-making. We developed a general modeling cancer framework with three capacity constraints along the cancer continuum (screening, diagnosis, and treatment). Then, we applied this framework to breast cancer healthcare constraints in Mexico and the United States.

3 - Effective and Efficient Staff in Emergency Departments

Hossein Hejazian, Harvard University, Cambridge, MA, United States, Soroush Saghafian

Emergency Departments (EDs) are complex, complicated, and always congested. Despite the extensive literature on ED operations, it is still an open question who constitutes an effective and/or efficient ED staff. This question is crucial in forming care teams to achieve the highest efficiency of the ED. We use a data set of ED visits in a leading US hospital to study the challenge of evaluating ED staff performance. Our aim is to propose a metric that combines the different aspects of a staff's productivity and the quality of delivered care. We investigate how various combinations of ED staff can impact operational efficiency and patient outcomes. Additionally, we analyze the potential peer effects among the ED staff. Our ultimate goal is to help our partner hospital attain higher efficiency levels by improving the formation of care teams (attending, nurses, residents, etc.) in their ED through a better understanding of the inherent trade-offs.

TC48

Summit - 440

Innovative Strategies for Efficient and Sustainable Systems

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yunlong Peng, University of Warwick- Warwick Business School, Coventry, United Kingdom

1 - Discriminatory Pricing in Queuing Systems: The Curse of Non-Transparency

Zihao Chen, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Xingyu Fu, Ying-Ju Chen

We study the pricing decision of a server within a queuing system. The server can differentiate between high-valuation consumers and low-valuation consumers and can implement discriminatory pricing. Surprisingly, we find that when congestion is at an intermediate level, price discrimination can harm the server's profitability. This unexpected outcome stems from the fact that discriminatory pricing results in a lack of transparency regarding prices, causing high-valuation consumers to distrust the server's promise to serve them exclusively. In other words, it is hard for the server to resist the temptation/opportunism to secretly include low-valuation consumers. As a result, high-valuation consumers' motivation to make purchases is dampened. To address these concerns, the server needs to lower prices to attract high-valuation customers, thereby reducing the server's profit margin. We demonstrate that the preemptive prioritization of high-valuation consumers can serve as a credible commitment mechanism and effectively address the non-transparency issue related to discriminatory pricing in queuing systems.

2 - Crowdsourcing Electric Mobility for Omni-Sharing Distributed Energy Resources

Wenqing Ai, University of Science and Technology Beijing, Beijing, China, People's Republic of, Tianhu Deng, Wei Qi

Ever-increasing coupling of energy and mobility sectors is underway in our cities. However, whether and how to use such coupling to optimize the portfolio and operations of urban energy assets has rarely been studied. We fill this gap by studying “omni-sharing,” which is a novel business model (beyond “peer-to-peer” energy sharing) that allows a distributed energy resource (DER) community to crowdsource electricity from shared electric vehicles (EVs). We analytically model two salient features of omni-sharing operations: the optimal payment to crowdsourced EV drivers, and the cost allocation among internal DERs. In doing so, we enrich the existing newsvendor cooperative game theory results by generalizing the newsvendor model with a variable underage cost structure, induced by the endogeneity of the energy-crowdsourcing cost. We prove (and quantify with data on residential energy consumption and ride-sharing markets) that omni-sharing can reduce the total storage capacity needed and the expected energy cost for the DER community. To account for intra-day energy demand fluctuation and supply uncertainty, we also develop a real-time dispatch model. Our analysis shows the value of omni-sharing as an additional channel of efficiently matching local energy supply and demand. Omni-sharing exerts mixed (but on balance positive) externalities: positive effects on ride-share drivers, and moderate and reducible negative effects on ride-share passengers and platforms. Moreover, omni-sharing enhances the urban grid’s self-sufficiency in terms of both peak load reduction and ramping mitigation. This paper strengthens our understanding of energy-mobility orchestration toward enabling smart cities.

3 - Harnessing Collective Wisdom: Crowd Judging in Enhancing Dispute Resolution on Two-Sided Platforms

Chong Zhang, Tilburg University, Tilburg, Netherlands, Zhonghong Kuang

Emerging digital platforms are introducing innovative mechanisms for resolving disputes by leveraging crowd judging as an alternative to traditional adjudication systems. This study investigates the effectiveness of crowd judging in improving dispute resolution on two-sided platforms, particularly in scenarios involving meticulous reviews and fastidious customers. We develop a game-theoretic model in which sellers can either present evidence proving their innocence or introduce misleading information to sway judges. By comparing conventional platform-based adjudication with the crowd judging mechanism, we aim to determine if the latter offers superior market efficiency and increased sellers' profits. Additionally, we explore the impact of information sharing among judges on the efficiency of the crowd judging process, assessing its potential to enhance decision-making accuracy.

4 - Assortment Personalization in A Subscription Model

Jiannan Xu, Robert H. Smith School of Business, University of Maryland, College Park, MD, United States, Gah-Yi Ban, Wedad Elmaghraby, Beril Toktay

We investigate the assortment personalization problem in a subscription model. Specifically, this paper focuses on a clothing rental subscription model where customers can rent clothes from a personalized closet of designer brands. We incorporate product fit uncertainty, i.e. the uncertainty surrounding how the customers will ``feel'' once they receive the garments and wear them into the subscription decision model, and consider nested constraints, i.e. that the firm’s decision is subject to the assortment cardinality constraint and the customer’s choice is subject to the subscription basket constraint, in the assortment personalization problem. We address how firms should select an optimal assortment considering (1) the product fit uncertainty and (2) the nested constraints to maximize expected lifetime revenue from a customer; we call this the doubly combinatorial assortment personalization problem. We show that this problem is NP-hard. Analytically, we derive optimal results for special cases and come up with near-optimal approximations for general cases with binding constraints. Computationally, our experiments on randomly generated instances demonstrate that our approximation algorithms, on average, yield expected revenues that are within 2.6% of the optimal expected revenue. Finally, we derive insights on how subscription service providers can optimize assortments by thinking products as functions of signal-to-noise ratios.

TC49

Summit - 441

Scheduling of Manufacturing and Service Processes

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Norbert Trautmann, University of Bern, FM Quantitative Methoden, Bern, 3012, Switzerland

1 - Online Quotation Under Delayed Customer Feedback and Offline Production Scheduling

Kelvin Ekun, University of Massachusetts, Amherst, MA, United States, Ana Muriel

While make-to-order (MTO) manufacturers provide order quotations upon request and may face penalties for not meeting the promised lead times, customers are typically offered a grace period over which to make the final decision to place their quoted order. As a result, the MTO firm faces an online lead time quotation problem with delayed feedback on order acceptance. The shorter the quoted lead time, the higher the probability of acceptance. Meanwhile, the production department will optimize and set the production schedule offline at regular intervals, given the current firm orders. We propose a primal-dual algorithm for this online quotation problem with delayed feedback and the periodic offline scheduling objective is to maximize total profit penalized by tardiness costs. Using simulation-based experiments, we show the performance of the proposed method relative to a hindsight oracle and show that it outperforms benchmarks used in practice.

2 - Flow Shops with Reentry: Total Completion TIME Optimal Schedules

Nicklas Klein, University of Bern, Bern, Switzerland, Michael Pinedo

Flow shops are widely studied machine environments in which all jobs must visit all machines in the same order. While conventional flow shops assume that each job traverses the shop only once, many industrial environments require jobs to pass through the shop multiple times before completion, i.e., after traversing the shop and completing its processing on the last machine, a job must return to the first machine and traverse the shop again until it has completed all its required loops. There are numerous applications, e.g., in semiconductor manufacturing, of such a setting, which is called a flow shop with reentry. The planning problem is to schedule all loops of all jobs while minimizing the total (weighted) completion time. We consider a reentrant flow shop with unit processing times and show that the Least Remaining Loops First (LRL) priority rule minimizes the total completion time. Furthermore, we show that the problem of minimizing the total weighted completion time is NP-hard. For this objective, we show that the Weighted Least Remaining Loops First (WLRL) priority rule has a worst-case performance ratio of about 1.2.

3 - A Hybrid Column Generation and Constraint Programming Approach for the Multi-Project Scheduling Problem with Local Resources

Maximilian Kolter, Technical University of Munich, Munich, Germany, Chris Beck, Rainer Kolisch, Martin Grunow

This research tackles the multi-mode resource-constrained multi-project scheduling problem with global and local resources. Global resources are shared among projects, and local resources belong to a specific project. While many heuristic algorithms exist for this problem, exact approaches for large-scale instances are missing. Hence, we propose an exact branch-and-price algorithm. Our algorithm is based on a reformulation of the classical discrete-time formulation. The reformulation allows the decomposition of the problem by projects and gives tighter linear relaxations than the classical formulation. However, the reformulated problem requires a huge number of variables. Thus, we solve the linear relaxations of the reformulated problem via column generation. The resulting pricing problems are NP-hard scheduling problems, which we solve via constraint programming. We validate the effectiveness of our algorithm benchmarking against commercial integer programming and constraint programming solvers.

4 - Workload Balancing in Integer Programming Models for Project Scheduling

Norbert Trautmann, University of Bern, Bern, Switzerland, Nina Ackermann, Tamara Bigler

A single-item or small-series production environment can be represented as a project consisting of a set of precedence-related activities that require time and resources to be performed. The resources represent, for example, production personnel. We present a mixed binary linear program for minimizing the makespan subject to workload balancing constraints.

TC50

Summit - 442

Stochastic Models for Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Vahid Sarhangian, University of Toronto, Toronto, ON, Canada

Co-Chair: Jangwon Park, University of Toronto, Toronto, ON, Canada

1 - Mitigating Nurse Shortage via Inter-Hospital Nurse Transfers: a Sample Robust Optimization Approach

Mengshi Lu, Purdue University, West Lafayette, IN, United States, Wei Liu, Tianchun Li, Pengyi Shi

We consider a nurse transfer problem that involves moving nurses from hospitals with surpluses to hospitals with shortages, thereby matching the demand and supply at each hospital. Specifically, we consider a fully connected network, allowing transfers between any two hospitals. Transferred nurses might need to stay at the destination hospital for a few days, known as secondment, to avoid long commutes. Additionally, the nurses need to receive advance notice of their transfers. However, a demand shift from the historical sample path—caused, for example, by the rapid spread of COVID-19 and the statewide distribution of hospitals—makes planning these transfers challenging. To address this, we adopt a sample robust optimization (SRO) approach. However, due to its complexity, this is challenging to solve. To simplify, we reformulate the problem into a more manageable linear optimization problem using linear decision rules. We show that the network design has a more significant impact on reducing total costs, and other related metrics including the transfer volumes and transfer distances compared to the secondment scenarios and the methods applied (SAA and SRO). Furthermore, our analysis reveals that longer secondments notably reduce these metrics as well. Moreover, we show that the SRO strategy is preferable in contexts where increases in nurse demand are anticipated, as it effectively minimizes the necessity for emergency transfers. This underscores the importance of considering a fully connected network design, secondments, and the SRO approach to enhance cost efficiency, minimize emergency transfers, and improve the system's responsiveness to demand fluctuations.

2 - Early Birds Versus Last-Minute Arrivals: Empirical Evidence and Theoretical Analysis of Arrival TIME Queueing Game

Xiangjie Zhao, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Yichuan Ding, Dongdong Ge, xiaoqing xie

In this study, we examine a scenario where each customer strategically selects their arrival times at a queue, aiming to minimize their total costs, which include waiting, tardiness, and earliness costs. Building on the mixed strategy framework characterized by Haviv (2013), which provides a closed-form arrival time distribution, our focus extends to the transient behavior of this system in presence of irrational customers, in which case it is not yet known whether an equilibrium exists or whether the system would converge to the equilibrium. We introduce a dynamic choice model to describe customers' behavior in a repeated game setting. We empirically validate this model using data from a large endoscopy clinic in China, operating on a first-come-first-served basis. The empirical result supports the efficacy of this dynamic choice model in capturing the adaptive strategies of patients concerning their arrival times. Leveraging this model, we prove that there exists a unique equilibrium and the system converges to this equilibrium under both the fluid model and the M/M/1 queue model. Furthermore, we formulate a dynamic control problem to characterize the optimal information sharing policy, suggesting that service providers may decrease social cost by deliberately modulating the communication of waiting time information to patients. Further examination of social cost unveils that increasing earliness cost can, counter-intuitively, detract from social cost at equilibrium.

3 - Mass Vaccination Scheduling: Trading off Infections, Throughput, and Overtime

Shanshan Luo, UBC Sauder School of Business, Vancouver, BC, Canada, Steven Shechter

Mass vaccination is essential for epidemic control, but long queues can increase infection risk. We focus on scheduling arrivals at mass vaccination sites to minimize a tri-objective function of infection risk, throughput, and overtime. Leveraging multi-modularity results from a related optimization problem, we construct a solution algorithm and find that our model-based policy significantly outperforms an equally-distributed, equally-spaced schedule. We also discuss managerial insights regarding the optimal schedule's structure and compare it to the well-known "dome-shaped" policies in appointment scheduling problems.

TC51

Summit - 443

Recent Advances in Data-Driven Decision-Making

Invited Session

MSOM: Service Operations

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Stefanos Poulidis, INSEAD, Fontainebleau, 77300, France

1 - Transfer Learning Through Directional Gradients

Vishal Gupta, University of Southern California, Los Angeles, CA, United States

We propose a novel decision-aware approach for multitask learning (data-pooling) tailored to low-data environments. Our approach is flexible and general purpose. The key idea is to leverage zeroth order directional gradients to estimate out-of-sample performance of particular pooled policies. For well-behaved policies, these directional gradient estimates are themselves differentiable and can be (approximately) optimized via SGD. Our method provably identifies best-in-class policies and performs well on real data with minimal tuning.

2 - Contextual Linear Optimization with Bandit Feedback

YICHUN HU, Cornell University, Ithaca, NY, United States, Nathan Kallus, Xiaojie Mao, Yanchen Wu

Contextual linear optimization (CLO) uses predictive observations to reduce uncertainty in random cost coefficients and improve performance. An example is stochastic shortest path with random edge costs. Existing work on CLO focuses on data with fully observed cost coefficient vectors, but in many applications we can only see the realized cost of a historical decision, that is, just one projection of the random cost coefficient vector. We study a class of algorithms for CLO with this so-called bandit feedback which we term induced empirical risk minimization (IERM), where we fit the predictive model to directly optimize downstream performance of the policy it induces. We show a fast-rate regret bound for IERM that allows for misspecified model classes and flexible choices of the optimization estimate, and we develop computationally tractable convex relaxations. We further demonstrate and compare the performance of different modeling choices numerically using a stochastic shortest path example, and provide practical insights from our empirical results.

3 - Quantifying the Cost of Learning in Queueing Systems

Wentao Weng, MIT, Cambridge, MA, United States, Daniel Freund, Thodoris Lykouris

Queueing systems are widely applicable stochastic models with use cases in communication networks, healthcare, service systems, etc. Although their optimal control has been extensively studied, most existing approaches assume perfect knowledge of the system parameters. This assumption rarely holds in practice where there is parameter uncertainty, thus motivating a recent line of work on bandit learning for queueing systems. This nascent stream of research focuses on the asymptotic performance of the proposed algorithms.

In this work, we argue that an asymptotic metric, which focuses on late-stage performance, is insufficient to capture the intrinsic statistical complexity of learning in queueing systems which typically occurs in the early stage. Instead, we propose the \emph{Cost of Learning in Queueing (CLQ)}, a new metric that quantifies the maximum increase in time-averaged queue length caused by parameter uncertainty. We characterize the \textsc{CLQ} of a single-queue multi-server system, and then extend these results to multi-queue multi-server systems and networks of queues. In establishing our results, we propose a unified analysis framework for \textsc{CLQ} that bridges Lyapunov and bandit analysis, provides guarantees for a wide range of algorithms, and could be of independent interest.

4 - Aligning Model Properties via Conformal Risk Control

William Overman, Stanford University, Stanford, CA, United States, Jacqueline Vallon, Mohsen Bayati

AI model alignment is crucial in modern machine learning due to the underspecified pipeline where numerous models with excellent test set metrics may not meet end-user requirements. Recent advances show that post-training model alignment via human feedback can address some of these challenges. However, these methods are often limited to settings like generative AI where humans can interpret model outputs and provide feedback. In traditional non-generative settings, where model outputs are numerical values or classes, detecting misalignment through single-sample outputs is more challenging.

We consider an alternative strategy. We propose interpreting model alignment through property testing, defining an aligned model as one belonging to a subset of functions that exhibit specific desired behaviors. We focus on post-processing a pre-trained model to better align with this subset using conformal risk control. Specifically, we develop a general procedure for converting queries for a given property to a collection of loss functions suitable for use in a conformal risk control algorithm. We prove a probabilistic guarantee that the resulting conformal interval around the model contains a function approximately satisfying the desired property. We exhibit our methodology on a collection of supervised learning datasets and for a variety of properties such as monotonicity and concavity. The general procedure is flexible and can be applied to a wide range of desired properties.

TC52

Summit - 444

Leveraging AI in Behavioral and Service Operations

Invited Session

MSOM: Service Operations

Chair: Paige Tsai, Harvard Business School, Somerville, United States

Co-Chair: Ryan Buell, Harvard Business School, Boston, MA, United States

1 - Dynamic Scripting in Call Center Operations

Michelle Kinch, Tuck School of Business at Dartmouth College, Hanover, NH, United States

Artificial intelligence is being rapidly deployed to change the face of customer service operations. In this study, we examine efficiency-service tradeoffs when AI is used to reinforce service script protocols used by call center agents.

2 - How Prediction Intervals Improve Human Algorithm Collaboration

Chantale Köster, University of Cologne, Cologne, Germany, Cedric Lehmann, Andreas Fuegener, Ulrich Thonemann

For managerial decision tasks, humans and algorithms often work together with the intention to combine their skills and thereby achieve complementary performance, that is, a higher performance than either party could achieve on their own. Data from practice and research suggests that algorithmic advice often improves human decisions, however, not beyond the algorithmic performance. Trust calibration due to missing collaboration mechanisms are seen as the main reason for unexploited complementary performance potential. One potential collaboration mechanism is to communicate algorithmic certainty. In this paper, we analyze how human decision making in algorithmically supported tasks is affected by the provision of prediction intervals. In a laboratory experiment, participants worked on a forecasting task in which they and the algorithmic advisor had complementary information. We show that prediction intervals are an effective mechanism that improve trust calibration. This way, decision makers rely more on accurate advice that comes with high certainty and less on inaccurate advice that comes with low certainty, leading to a higher complementary performance outperforming both - humans and algorithms - working alone. Our results contribute to a better understanding of how humans and algorithms can achieve complementary performance. We suggest that managers consider the provision of prediction intervals in algorithmically supported decision making, since they lead decision makers to efficiently use algorithmic advice and achieve complementary performance.

3 - Unpacking the Dynamics of Customer Experience in AI-Infused Service Workflows

Cecilia Ying, Queen’s University, Markham, ON, Canada

As Generative AI transforms customer service, understanding the interplay between customers and agents is crucial. This study investigates the key drivers of negative customer experiences before and after AI integration, shedding light on the contrasting dynamics between the customers' interaction with the human-agent verses the AI-agent.

4 - One Versus Many: How Users Integrate Advice from Multiple Algorithms

Anna Taudien, University of Cologne, Cologne, Germany

Today, humans frequently receive algorithmic advice for prediction tasks in various fields such as healthcare, logistics, or finance. So far, existing research has primarily focused on collaboration settings where a user is advised by a single AI. However, the increasing presence of AI suggests that human-AI teams will likely involve multiple AIs in the near future. Yet, little is known on how humans will integrate advice from multiple algorithmic advisors into their judgment. In two between-subjects experiments ($n = 401$), we studied this advice-setting by having subjects perform surgery-duration predictions with varying numbers of algorithmic advisors. We find that users increase their weight on algorithmic advice when being advised by more than one algorithm. This not only occurs when the additional advice is beneficial (i.e., users can realize performance gains) but also when it is harmful (i.e., users impair their own performance). Hence, users do not seem to rationally include the additional algorithmic advice by weighing the different sources against each other. Instead, they appear to increase their reliance on algorithmic advice simply because more options are available. This behavior aligns with the bias of naïve diversification which describes the tendency of overly diversifying between different options. Our results contribute to the literature on AI-advised settings and have implications for organizations aiming to implement multiple algorithms as team members for humans.

5 - Learning about learning: Encouraging Algorithm Alignment through Progressive Deployment

Maya Balakrishnan, Harvard Business School, Boston, MA, United States, Bryce Mclaughlin

Across industries, managers’ decision-making processes evolve as their firms adopt new decision support systems (such as prediction algorithms). When a new algorithm is introduced to a user, the human’s responses to the algorithm are not static. The user’s initial interactions with the algorithm will drive how they learn about the algorithm and its performance and shape how they continue to use the algorithm in the future. At the same time, the algorithm needs to understand how users will respond to its deployment in order to effectively present its recommendations to achieve human-algorithm alignment. We propose a model by which a new algorithm is progressively deployed to heterogeneous users where users iteratively learn about the algorithm while the algorithm simultaneously learns about users and their response to the algorithm. We showcase our progressive algorithm deployment process in a series of simulated experiments and compare its performance against non-dynamic deployment processes, highlighting the importance of carefully managing not just what algorithm is deployed but how it is introduced.

TC53

Summit - 445

Sustainability in the Food Supply Chain

Invited Session

MSOM: Sustainable Operations

Chair: Sridhar Seshadri, University of Illinois, Champaign, IL, United States

Co-Chair: Sandeep Srinivas, University of Illinois at Urbana-Champaign, Champaign, IL, 61820

1 - Impact of Blockchain Technology on Food Safety

Sandeep Srinivas, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Several regulatory bodies have underscored the significance of traceability in the food supply chain to safeguard consumer health. Numerous food and beverage companies have recently adopted blockchain technology to enhance supply chain traceability. This study empirically examines the effectiveness of blockchain traceability in improving food safety outcomes, specifically food recalls. We find that adopting blockchain traceability is associated with reducing recall duration for future recalls by firms. Additionally, we analyze the impact of blockchain traceability adoption on the number of recalls and recall volumes.

2 - Markdowns to Manage Food Waste and Its Impact on Retail Brand Image

Arzum Akkas, University of Massachusetts Amherst, Amherst, MA, United States

While marking down surplus food is an effective strategy to reduce food waste in retail, we discovered that many retailers, especially in the US, are reluctant to adopt this approach due to concerns about their brand image. Through both in-store and online experiments, we investigated whether markdowns have a negative impact on retailers' brand perception. Our findings revealed that they do not. Additionally, we found that the way markdowns are presented is important. Our online experiments also revealed geographical variations across the UK, US, Germany, France, and South Africa.

3 - Achieving Resilience with Redundancy in Food Bank Operations

Pelin Pekgun, Wake Forest University, Winston-Salem, NC, United States, Yingru Han, Luv Sharma, Olga Perdikaki

Food banks are non-profit organizations that collect food from sources such as retailers, manufacturers, and farmers, and distribute to the needy population. In this study, we investigate the role of redundancy-building strategies in affecting a food bank's level of resilience in food distribution in the case of a major disruption.

4 - Retail Food Donations and Operational Efficiency

John Lowrey, Northeastern University, Boston, MA, United States, Tim Richards, Anne Byrne

Retail unsaleables are often linked to an opportunity cost, since the resources used to produce leftover food could have otherwise been used more productively. Leftover food at retail has many potential uses, however, including donations to food banks. We examine the relationship between retail food donations and operational efficiency using a sample of 1,581 store-years across 14 major banners that donated to two major Feeding America member food banks from 2015-2021. We use stochastic frontier analysis and account for the endogeneity of key productive inputs using a control function framework. We find that food donations increased operational efficiency for stores located in more deprived areas. We link these results to potential differences in unsaleables management: stores in more deprived areas rely on donations to avoid cost, while stores in less deprived areas use donations to maximize quality (and profit). The cost-push vs. profit-pull difference may extend to other aspects of retail operations, beyond unsaleables management.

TC54

Summit - 446

Empirical Research on Operational Efficiency and Performance

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: So Yeon CHUN, INSEAD, France, Singapore

Co-Chair: Priscilla Rodriguez, INSEAD, Fontainebleau, France

1 - Algorithmic Decision Making in Healthcare: Role of Non-Clinical Call Handlers

Deepanshi Bhardwaj, UCL School of Management, London, United Kingdom, Bilal Gokpinar

Using data from NHS 111 and its Computer Decision system (Pathways), this paper aims to understand the extend to which service agents rely on and trust algorithms. To meet the growing demand of health services, the NHS uses a helpline number 111 to deal with urgent but non-life-threatening emergencies. These calls are handled by non-clinical staff using a Computer Decision Support System (NHS Pathways). In the precursor to NHS 111 (NHS Direct), the Decision Support System was designed to support, rather than replace, clinical decision-making. However, in case of NHS 111, there is an assumption (by NHS 111 service providers and by the technology developers) that clinical decision-making is primarily a function of the technology (provided that the call handler correctly identifies the appropriate computer algorithm). We identify possible errors that occur by over-reliance and under-reliance on Algorithmic Decision Support Tool and how the errors impact the Healthcare System.

2 - Performance Implications of Supply Chain Turnover

Kevin Mayo, Washington State University, Pullman, WA, United States, Christopher Chen, George Ball, Kurt Bretthauer

We examine the operational performance consequences of supply chain turnover. Analyzing a panel of buyer-supplier turnovers in the consumer goods industries, increasing turnover increases the consumer complaints, signaling a decrease in product quality, offsetting the decreased costs associated with turnover. We find that costs increase and quality remains the same when the turnover is due to the supplier exiting the market. We investigate the effects of distance and economic development on the supplier portfolio as mechanisms to explain the result.

3 - Going the extra mile? Workplace behaviors and performance outcomes.

Priscilla Rodriguez, INSEAD, Fontainebleau, France, So Yeon CHUN, Jurgen Mihm

In today's competitive corporate landscape, understanding the relationship between workplace behaviors and performance outcomes is crucial. We leverage privileged access to email and meeting data spanning over 12 thousand employees within a firm to understand the dynamics of employee behavior. Specifically, we focus on what we call "extra mile behaviors," characterized by employees' willingness to work extended hours and weekends. We analyze these behaviors in light of performance outcomes, focusing on differences in effects by employee characteristics and the potential for employee burnout.

4 - How Do Loyalty Programs Affect Multichannel Bookings? An Empirical Investigation

Shuaishuai Yang, UCL School of Management, London, United Kingdom, Antoine Feylessoufi, Ersin Korpeoglu, Lina Song

With the increased market power of online travel agencies (OTA), the hotel industry faces a new and important trade-off when working with OTAs: Potentially increasing sales at the expense of sizeable commission fees. In response, major hotel groups turn to loyalty programs, giving discounts to loyalty members for booking through their direct channels. In this paper, we investigate how successful loyalty discounts are in their goal to sway customers from OTAs to direct channels. Using a unique dataset obtained from a major hotel group, we empirically investigate how loyalty discounts affect customers' booking channel choices. We utilize a quasi-natural experiment based on an exogenous change of the loyalty discount in a subset of hotels in our dataset to construct a difference-in-difference model. We find that a larger loyalty discount raises bookings from the direct channel with loyalty membership,lowers bookings from the direct channel without loyalty membership, and has no significant effect on OTA bookings. This suggests that a substitution effect only occurs within the direct channel and, surprisingly, not cross-channels. Our results indicate that discount-based loyalty memberships do not have the intended effect of swaying customers from OTAs toward direct channels. Instead, directchannel customers utilize the loyalty program more, which may end up lowering hotel revenues. Therefore, our study suggests that the hotel industry should carefully assess the effectiveness and the design of their loyalty programs by factoring in their multichannel setting.

TC55

Summit - 447

Use-inspired analytics and models tackling Human Trafficking

Invited Session

Public Sector OR

Chair: Felipe Aros-Vera, Ohio University, Athens, OH, United States

1 - Jennifer Ppt

2 - Optimum Allocation of Rehabilitation and Reintegration Services to Support Victims of Human Trafficking

Riad Al Hasan Abir, Ohio University, Athens, OH, United States, Felipe Aros-Vera

Human trafficking (HT) is a grave violation of human rights with far-reaching impacts, touching lives in every corner of the world. Despite the existence of government and non-profit rehabilitation services, HT-affected individuals often miss out due to improper resource allocation. To address this issue, we propose an optimization model that efficiently allocates resources to rehabilitate and reintegrate HT-affected individuals where those resources are most needed. Our strategy uses a Mixed Integer Linear Programming (MILP) model to maximize the net societal value (NSV) gained from offering support services while considering the three stages of the healing path for HT-affected people, including victim, survivor, and thriver. This model determines the optimal type, quantity, and location of services, while also integrating HT risk scores that account for the risk of HT in those areas. Our model's efficacy is demonstrated in a case study in Ohio, that allocates housing, detoxification, and food services across the state's eighty-eight counties and the three stages of the healing path of HT-affected individuals. Through Monte Carlo Simulation in the solution approach, uncertain demand is accounted for, leading to improved NSV under such conditions. Moreover, we illustrate the impact of an increased budget, showcasing the reach and allocation possibilities of extended services. Our work aims to support decision makers in efficiently allocating resources to rehabilitate and reintegrate HT-affected individuals effectively.

3 - Machine Learning Solutions in Identifying Potential Human Trafficking Recruitment During Humanitarian Disasters: Case Study of Ukraine

Amir Jamali, Worcester Polytechnic Institute, Worcester, MA, United States, Navid Ziaei, Renata Konrad

Humanitarian disasters increase vulnerabilities, as well as opportunities for exploitation, particularly of displaced people and refugees. Since the 2022 full-scale invasion of Ukraine, millions have been displaced, many unable to find licit employment and stable housing, putting them at risk of exploitation and human trafficking.

Online technologies such as home hosting sites and chat groups on social media are used to seek housing and employment, offering a lifeline to many but also facilitating exploitation.

We developed a machine learning model that uses natural language processing to analyze Telegram data, identifying potentially suspicious recruitment activities. This model classifies messages to pinpoint potential human trafficking communications. To handle multilingual data, we translated non-English messages into English, standardized non-text elements such as emojis, and tailored our methodology to accurately distinguish between suspicious and non-suspicious interactions.

To overcome data imbalance, we refined our labeling process and reduced the volume of non-suspicious data, enhancing the model's

performance and reliability. This approach allows us to identify vulnerabilities more effectively and supports specialists in directing prevention and rehabilitation efforts where they are most needed.
Our ongoing evaluation ensures the model's reliability in real-world applications. This automated tool saves considerable time in manual data screening, allowing NGOs and law enforcement to concentrate on intervention and support. The model not only contributes significantly to combating human trafficking in Ukraine but also offers a scalable solution adaptable to other regions, enhancing global anti-trafficking efforts.

4 - Modeling Human Trafficking Illicit Supply Networks

Felipe Aros-Vera, Ohio University, Athens, OH, United States, Stacie Petter, Gisela Bichler

With 50 million people affected around the world, human trafficking can (and should) be considered a humanitarian crisis. Illicit supply networks facilitate exploitation and threaten legitimate economic systems. Borrowing network science and criminology concepts, this presentation identifies key behavioral and networked characteristics that influence human trafficking within illicit supply networks.

5 - Improving Network Interdiction Models for Combating Labor Trafficking: Insights and Limitations

Priscila de Azevedo Drummond, Northeastern University, Boston, MA, United States, Arezoo Jafari, Kayse Lee Maass

Labor trafficking research faces considerable challenges due to data scarcity, surpassing even those encountered in sex trafficking studies. Although Network Interdiction Problem (NIP) models have been used to tackle crimes such as nuclear smuggling and drug trafficking, their application to labor trafficking is still in its early stages. Incomplete knowledge of the structural dynamics of trafficking operations and the extent of its prevalence influences this lack of publishing.
Our research focuses on the challenges of using NIP models to help reduce labor exploitation and proposes strategies to overcome them. However, assumptions and network structures chosen greatly impact the interpretative capabilities of these models. Through an interdisciplinary team, we build off expert knowledge to draw a more nuanced understanding of the models' relevance to the field. By exploring different models' assumptions, we found new opportunities for NIP model applications. Our findings have the potential to provide practical insights into the dynamics of labor trafficking and the effectiveness of anti-trafficking interventions.

6 - Does Deregulation Impact Firms' ESG Policy Compliance?

Jennifer Nowicki, Pennsylvania State University, University Park, PA, United States, Brian Lee, Kevin Linderman

The mining of conflict minerals is associated with regional violence, forced labor, and environmental devastation, which poses risks to firms' supply chains and therefore increases the potential for operational disruptions. As offshoring proliferated in the 2000s, these conflict regions have become increasingly entangled in firms' upstream supply chains, and the 2010 Dodd Frank Act requires firms to investigate and monitor their operations for the purpose of reducing risk. This legislation includes a conflict minerals provision to account for humanitarian risks imparted through conflict mining, which funds armed groups in the region. To comply with the legislation, firms must determine whether they are sourcing 3TG materials from a designated conflict region and report their findings to the SEC, however this provision was partially dismantled in 2017, providing an opportunity to observe firms' compliant behavior for both policy enactment and rollback. Through analyzing SEC documents via natural language processing, we explore how firms responded to the partial deregulation of the Dodd Frank Act's conflict minerals provision over a 10-year span, and evaluate firms' risk management systems through a due diligence framework. We find significant contingencies that impact firms' compliance behavior post-deregulation through a series of empirical analyses, and pair these findings with insights from industry practitioners. We then conclude with a discussion on a proposed compliance model for ESG-related policies, an increasingly relevant topic as ESG legislation propagates worldwide.

TC56

Summit - 448

Advancements in Quantum Computing and Collaboration

Contributed Session

Chair: Jiaqi Leng, University of Maryland, College Park, MD, United States

1 - The co-evolution dynamics of multi-agent collaboration network in the emerging industry: A stochastic actor-oriented model in the field of quantum computing

Yingjie Ma, Zhejiang University, Hangzhou, China, People's Republic of, Xuansheng Wang, Qihan Yang, Zhiyang Liu, Lifan Chen

The realization of quantum supremacy highlights the potential of quantum computing to make a revolutionary difference. However, as an emerging industry, the quantum computing field has yet to establish a dominant design, leading to uncertain schemes of industrialization and commercialization. Drawing on the perspective of complex network and social embeddedness theory, this study explores the dynamics of collaborative innovation among key stakeholders and changes of research priorities in the field of quantum computing. Focusing on China and the United States as representative countries, this paper collects bibliographic data between 2001 and 2023 to construct inter-institutional collaboration networks. The present paper conducts two studies. In Study 1, we select a set of complex network measurements to characterize the evolution of topological structures of collaboration network in the quantum computing field. Additionally, we build dynamic topic models of the leading countries using BERTopic technique. Building upon the results in Study 1, Study 2 uses stochastic actor-oriented model (SAOM) to uncover the underlying mechanisms driving the co-evolution of collaboration network structures and innovation output of different agents, accounting for structural effects and agent attributes. The findings are: (1) Quantum computing is currently situated in a phase of fundamental research, marked by the dominance of academia, the increasing focus on market demands, and the consistent support from government; (2) Geographical, institutional, organizational, cognitive, and social proximity all influence the selection of collaborators, albeit variances at different stages; (3) The whole ecosystem remains unstable. Theoretical and managerial implications are discussed.

2 - Challenging the Quantum Supremacy Claim Using Learning-to-Anneal Algorithm

Xiao-Yang Liu, Columbia University, New York, NY, United States

Google's "quantum supremacy" announcement has received broad questions from academia and industry due to the debatable estimate of 10, 000 years' running time for the classical simulation task on the Summit supercomputer. Has "quantum supremacy" already come? Or will it come in one or two decades later? To avoid hasty advertisements of "quantum supremacy" by tech giants or quantum startups and eliminate the cost of dedicating a team to the classical simulation task, we take an open-source approach to maintain a trustable benchmark performance. We propose a novel learning-to-anneal algorithm, inspired by the simulated annealing algorithm and reinforcement learning algorithm, for the classical simulation of quantum circuits. We demonstrate its great potential by reporting an estimated simulation time of less than 4 days. Specifically, we formulate the classical simulation task as a tensor network contraction ordering problem, which is NP-hard by reducing to the problem of finding the ground state of Ising spin-glass model. We open-source our parallel gym environments and benchmarks and organize academic competitions to maintain the best performance of classical computers.

3 - The impact of collaboration networks on high-level scholars' innovation performance in the field of quantum information technology

baohuan zhou, University of Science and Technology of China, hefei, China, People's Republic of, Liang Liang

As a critical part of modern physics, quantum information technology is a field with thresholds, and analysis needs to be performed on high-level scholars. However, few studies have examined high-level scholars' innovation performance in the field from the perspective of collaboration networks. This article attempts to use methods such as bibliometrics, collaboration networks, and machine learning to fill these gaps. We constructed collaboration networks through two dimensions: scholars' egocentric networks and team networks. The effects were examined by considering two levels: innovation quantity and innovation quality. This study examines the connection between collaboration networks and high-level scholars' innovation performance. Meanwhile, 6,587 high-level scholars extracted from 202,614 papers between 2000 and 2020 in the field of quantum information technology were used for analysis. The results of the analysis show new features of the collaboration effect among many high-level scholars in the field of quantum information technology. In terms of scholars' egocentric networks, occupying an intermediary or structural hole location positively affects scholars' innovation performance. From the team networks perspective, there is an inverted U-shaped relationship between team size and high-level scholars' innovation quality. Team size should be approximately 50~60 scholars and be determined in ranges depending on goals.

4 - Using Neutral Atom-based Quantum Computing for Solving Hard Combinatorial Problems

Yassine NAGHMOUCHI, PASQAL, Massy, France, Yassine NAGHMOUCHI

Neutral atom technology has steadily demonstrated significant theoretical and experimental advancements, positioning itself as a front-runner platform for running quantum algorithms. One unique advantage of this technology lies in the ability to reconfigure the geometry of the qubit register, from shot to shot. This unique feature makes possible the native embedding of graph-structured problems at the hardware level, with profound consequences for the resolution of complex optimization tasks. By driving qubits, one can generate processed quantum states which retain graph complex properties. These states can then be leveraged to offer direct solutions to problems or as resources in hybrid quantum-classical schemes. In this talk, we'll present in detail how neutral atom-based quantum computing can tackle complex combinatorial tasks on solving graph-related optimization problems.

5 - Quantum-inspired algorithms for nonconvex quadratic programming

Jiaqi Leng, University of Maryland, College Park, MD, United States, Samuel Kushnir, Yuxiang Peng, Xiaodi Wu

Quantum algorithms have drawn increasing attention in the operations research community due to their promising performance for hard optimization problems. However, in the near future, the application of quantum algorithms to large-scale problems is still limited by the scalability of quantum computers. Recently, quantum-inspired classical algorithms have emerged as an active area of research because they preserve a quantum-like advantage while being fully executable on classical hardware. In this talk, we will introduce a family of quantum-inspired algorithms for quadratic programming, a fundamental task in nonlinear programming. Formulated as Hamiltonian dynamical systems inspired by quantum mechanics, these algorithms exhibit a desirable tunneling effect that is unseen in conventional first- and second-order methods. Local convergence results are established for these algorithms, and empirical results show that the output solutions are often globally optimal.

TC57

Summit - Terrace Suite 1

Causal Inference

Invited Session

Health Applications Society

Chair: Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

Co-Chair: Raaz Dwivedi, Cornell Tech, Cornell ORIE, New York City, NY, United States

1 - Decision-Centric Causal Structure Learning: An Algorithm of Data-Driven Covariate Adjustment

Chandler Squires, Carnegie Mellon University, Pittsburgh, PA, United States, Davin Choo, Arnab Bhattacharyya, David Sontag

When learning a causal model of a system, a key motivation is the use of that model for downstream decision-making. In this talk, I will take a decision-centric perspective on causal structure learning, focused on a simple setting that is amenable to careful statistical analysis. In particular, we study causal effect estimation via covariate adjustment, when the causal graph is unknown, all variables are discrete, and the non-descendants of treatment are given.

We propose an algorithm which searches for a data-dependent "approximate" adjustment set via conditional independence testing, and analyze the bias-variance tradeoff entailed by this procedure. We prove matching upper and lower bounds on omitted confounding bias in terms of small violations of conditional independence. Further, we provide a finite-sample bound on the complexity of correctly selecting an "approximate" adjustment set and of estimating the resulting adjustment functional, using results from the property testing literature.

We demonstrate our algorithm on synthetic and real-world data, outperforming methods which ignore structure learning or which perform structure learning separately from causal effect estimation. I conclude with some open questions at the intersection of structure learning and causal effect estimation.

2 - Doubly Adaptive Algorithms for Threshold Identification in Multi-Armed Bandits

Brian Cho, Cornell University, New York, NY, United States, Nathan Kallus, Kyra Gan

In this work, we investigate the thresholding bandit problem (TBP) and its variants. In TBP with W arms, a learner is given a threshold vector, with the goal of identifying the arms whose means are above their corresponding thresholds as quickly as possible. Prior works for TBP mostly focus sampling schemes for well-studied parametric bandit instances, with stopping rules based on loose concentration inequalities. In our work, we provide both adaptive sampling schemes and adaptive stopping rules for TBP and its variants based on anytime valid testing procedures, resulting in doubly adaptive algorithms for both parametric and nonparametric bandit instances. In particular, our anytime valid confidence testing procedure adapts to the optimal sequential test for the realized bandit instance, achieving instance-wise optimality when paired with the correct sampling scheme. We first illustrate the power of adaptive stopping rules in achieving instance-dependent lower bound in TBP. Next, we focus on a specific variant of TBP where the goal is to promptly identify one arm above its corresponding threshold. We introduce an adaptive sampling scheme and show that when paired with our proposed stopping rule, it achieves the instance-dependent asymptotic lower bound, showing the benefit of double adaptivity. To illustrate the practicality of our proposed method with finite samples, we conduct numerical experiments that highlight the superior performance of our methods compared to existing state-of-the-art algorithms.

3 - Learning Counterfactual Distributions via Kernel Nearest Neighbors

Kyuseong Choi, Cornell University, New York City, NY, United States, Jacob Feitelberg, Anish Agarwal, Raaz Dwivedi

Consider a setting with many units (individuals, cohorts, geographic locations) and multiple outcomes (treatment, time, items). For each unit and outcome, there is a multivariate distribution associated with it that we would like to learn. Challenge here is in real world settings, most cells are missing as is the case for recommender systems, or causal panel data settings with staggered adoption where a sub-collection of units get treated after a period of time. Further, for the cells that we do observe, we have only access to finite number of samples from underlying distribution rather than the distribution itself. We model this problem as a distributional matrix completion problem, and introduce kernel nearest neighbors algorithm to learn counterfactual distribution of missing cell. Under latent factor model on kernel mean embeddings of distributions and regularity conditions, kernel nearest neighbors enjoy rate guarantees when data is not missing at random and when positivity constraints are violated. We also illustrate how more than one samples per cell can effectively make nearest neighbors robust to heterogeneous data by constructing unbiased estimates of dissimilarity measures, which is otherwise not possible with only one sample.

4 - A New Perspective on Synthetic Controls

Yujin Jeong, Stanford University, Stanford, CA, United States, Dominik Rothenhäusler

In statistics and machine learning, we often want to quantify uncertainty and frame optimality with respect to sampling uncertainty. However, if we combine evidence from different data sets, sampling uncertainty might be lower order than the distribution shift between the data sets. This raises the question of how to optimally estimate in a data fusion setting. To address this issue, we model distributional shifts as a superposition of numerous random changes. We then develop tools for measuring the similarity between randomly perturbed distributions, estimating parameters and predicting outcomes for partially observed target distributions. Interestingly, these tools share a close connection to synthetic controls. Our framework provides a new language for distributional shifts and offers a fresh perspective on synthetic controls. Finally, we apply the framework to several real-world data sets and discuss diagnostics to evaluate the fit of the distributional uncertainty model.

5 - Counterfactually Comparing Abstaining Classifiers

Yo Joong Choe, University of Chicago, Chicago, IL, United States, Aditya Gangrade, Aaditya Ramdas

Abstaining classifiers have the option to abstain from making predictions on inputs that they are unsure about. These classifiers are becoming increasingly popular in high-stakes decision-making problems, as they can withhold uncertain predictions to improve their reliability and safety. When evaluating black-box abstaining classifier(s), however, we lack a principled approach that accounts for what the classifier would have predicted on its abstentions. These missing predictions matter when they can eventually be utilized, either directly or as a backup option in a failure mode. In this paper, we introduce a novel approach and perspective to the problem of evaluating and comparing abstaining classifiers by treating abstentions as missing data. Our evaluation approach is centered around defining the counterfactual score of an abstaining classifier, defined as the expected performance of the classifier had it not been allowed to abstain. We specify the conditions under which the counterfactual score is identifiable: if the abstentions are stochastic, and if the evaluation data is independent of the training data (ensuring that the predictions are missing at random), then the score is identifiable. Note that, if abstentions are deterministic, then the score is unidentifiable because the classifier can perform arbitrarily poorly on its abstentions. Leveraging tools from observational causal inference, we then develop nonparametric and doubly robust methods to efficiently estimate this quantity under identification. Our approach is examined in both simulated and real data experiments.

TC58

Summit - Terrace Suite 2

Clinical Decision Making for Chronic Disease

Invited Session

Health Applications Society

Chair: Hyojung Kang, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Co-Chair: Jennifer Lobo, University of Virginia, Charlottesville, VA, United States

1 - Clustering Antidiabetic Medication Usage Patterns Using TIME Series Techniques

Peng Zhang, University of Illinois Urbana-Champaign, Urbana, IL, United States, Jennifer Lobo, Hyojung Kang, Min-Woong Sohn

Medication adherence in patients with type 2 diabetes has been suboptimal. Previous research has examined medication usage at the population level and revealed trends at specific treatment stages such as initiation and intensification. This study aims to develop a clustering model to categorize patients' medication usage patterns based on their longitudinal medication history. We focused on changes in medication usage patterns during the first year following metformin initiation. Dynamic time warping was employed to better evaluate the distances between different medication usage sequences. We examined the characteristics of clusters in terms of socioeconomic factors, as well as outcome measures, to understand the underlying disparities within diverse medication usage patterns and their impacts. The findings of this study will help identify high-risk patients for medication non-adherence and inform targeted patient management strategies.

2 - Machine Learning in the Emergency Department when the Class of Interest Adapts to System Congestion

M Gabriela Sava, Bowling Green State University, Bowling Green, OH, United States, Jerrold May, Ronald Pirrallo

Diabetes screening of Emergency Department/Urgent Care patients can proactively improve health outcomes, but it is uneconomic to screen all such patients. We present a data-driven analytical approach that could assist physicians with the diabetes screening decision. The approach is capable of selecting the most appropriate statistical model as resource availability and the patient's historical frequency of utilization change over time. Our findings show that when testing resources are more constrained, the predictive accuracy is greater for frequent users and decreases with patient visit frequency. Conversely, when testing resources are more available, the predictive accuracy decreases as patient visit frequency increases.

3 - Human-AI Interaction with Contextual Recourse Bandits

Ruijiang Gao, University of Texas at Austin, AUSTIN, TX, United States, Junyu Cao, Esmaeil Keyvanshokooh

This paper introduces a new paradigm called the \textit{Contextual Recourse Bandits} in online learning. The objective of this paradigm is to optimize outcomes by dynamically recommending actions and feature modifications. Drawing inspiration from practical healthcare scenarios where patients' features can be modified to qualify for more effective treatment options, we extend the traditional contextual bandit setup. Our contextual recourse bandit not only selects an action based on the current context but also suggests modifications to the context as recourses. These modified contexts have the potential to improve the outcome. To address this, we propose the Recourse Linear Upper Confidence Bound (RLinUCB), an innovative algorithm that jointly manages recourse and action recommendations, thereby balancing the exploration-exploitation trade-off. We provide theoretical guarantees demonstrating that RLinUCB achieves sublinear recourse regret under mild conditions. Furthermore, we offer a robust variant that accounts for noisy implementations of recourses. Empirical evaluations further validate the superior performance of RLinUCB compared to existing online contextual bandit algorithms. This work sheds light on a new direction in the contextual bandit literature that holds promise for advancing personalized healthcare among other fields.

4 - Comparison of Pancreatic Cyst Management Algorithms

Jennifer Lobo, University of Virginia, Charlottesville, VA, United States, Nasma Al-Humadi, Bryan Sauer, Ross Buerlein

Numerous guidelines exist for the management of pancreatic cysts that differ significantly in terms of their thresholds for imaging, surveillance, and surgery. We present a microsimulation model to compare outcomes for a cohort of patients managed with each guideline. Outcomes include mortality related to pancreatic cyst management, all-cause mortality, missed cancers, number of surgeries, number of imaging studies, cumulative cost, and quality-adjusted life years.

TC59

Summit - Ballroom 1

APS PhD Student Showcase 2

Award Session

Applied Probability Society

Chair: Daniel Russo, Columbia University, New York, NY, United States

Co-Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

TC60

Summit - Ballroom 2

Transportation Network Design

Invited Session

The Practice Section of INFORMS

Chair: Chun Ye, Amazon.com, Seattle, WA, United States

1 - Ecommerce Delivery Speed Through a Queuing Equilibrium Perspective

Amitabh Sinha, Amazon, Seattle, WA, United States

Leveraging simple models and learnings from Amazon's Regionalization initiative, we will examine the role of capacity allocation in ensuring fast delivery speed. We will develop a simple queuing model and use it to illustrate the impact of capacity imbalances on delivery speed, and build on that to develop principles for network design.

2 - Data Driven Network Design

Chun Ye, Amazon.com, Seattle, WA, United States, Sukanya Kudva, Daniel Ulch

We study the problem of optimizing network connectivity by deciding which pair of facilities to connect via truck routes. We consider a model that jointly optimizes connectivity, truck scheduling, and package flow using a deterministic MILP. However, the model's simplifying assumptions about truck scheduling and package flow often don't match real-world operations, which leads to suboptimal connectivity decisions. To address this, we discuss a method to refine these assumptions and improve the model's decision.

3 - Translating Science into Business Impact

Shahbaaz Mubeen Mamadapur, Amazon, Seattle, WA, United States

Business leaders need robust tools to help them make the right decisions. However, many a times, science teams are unable to get business leaders/teams to adopt a suggested approach. In this talk, we discuss how we leverage scientific tools and build scalable OR products within Amazon. We will go through how basic software engineering architectures and principles enable us to decouple the problems and develop generalized products to solve network design problems across multiple geographies and programs. Finally, we walk through specific use cases where the generalized products empowered business leaders to make go/no-go investment decisions and evaluate alternate network design choices.

4 - The Fulfillment Regionalization Problem

Nidhima Grover, Georgia Institute of Technology, Atlanta, GA, United States, Xiaoyan Si, Alejandro Toriello

In today's retailing formats, retailers can choose which inventory location or Fulfillment Center (FC) to fulfill from, bringing opportunities of inventory pooling and providing a much broader product selection. The fulfillment decision, although complex, can bring financial gains when optimized for resources and operating costs. With the unprecedented growth of the retail industry, companies now have the opportunity to strategically divide their fulfillment networks into regional networks. Such a method, called regionalization, simplifies the fulfillment decision, is scalable, and has helped to increase fulfillment speed while lower cost at Amazon. The region definitions affect speed of delivery, cost to serve, and selection of products offered, which influences customer experience. We present a heuristic approach on how to partition the national network into regions that fulfill customer demand predominantly from a set of pre-assigned FCs, and benchmark solution quality against lower bound models.

TC61

Summit - Ballroom 3

Healthcare - Radiotherapy Cancer Treatment Applications

Invited Session

The Practice Section of INFORMS

Chair: Masoud Zarepisheh, Memorial Sloan Kettering, New York, NY, United States

1 - Portpy: Opensource Package for Planning and Optimization for Radiation Therapy in Python

Gourav Jhanwar, Memorial Sloan Kettering Cancer Center, New York, NY, United States, Mojtaba Tefagh, Qijie Huang, Vicki Trier Taasti, Seppo Tuomaala, Saad Nadeem, Masoud Zarepisheh

We have developed PortPy (Planning and Optimization for Radiation Therapy in Python), an innovative opensource software package designed to expedite the research and clinical translation of novel treatment planning techniques in radiation therapy. PortPy addresses the limitations of existing packages by offering several key features: 1) Interoperability with Commercial Systems, allowing for the objective evaluation of new techniques in comparison to the current clinical practices. 2) Inclusion of Benchmark Dataset, currently 50 lung patients, complete with clinical plans and all necessary data for treatment planning optimization. 3) Advanced Benchmark Algorithms, capable of identifying "global" optimal solutions, using computationally intensive mixed integer programming, for various non-convex optimization problems encountered in treatment planning

2 - Utilizing Data Compression and Dimensionality Reduction to Combat the Curse of Dimensionality in Radiotherapy Cancer Treatment

Masoud Zarepisheh, Memorial Sloan Kettering, New York, NY, United States

This study aims to leverage the dimensionality reduction tools to improve the speed and quality of radiotherapy planning. The dimensionality reduction has a rich history in statistics and has recently re-emerged as a powerful tool to deal with the problems arising in the fields of big-data and machine learning. The optimization problems arising in radiotherapy also suffer from the curse of dimensionality (e.g., many machine parameters, voxels). We show how matrix/tensor decomposition, in particular low-rank plus sparse decomposition, and wavelets can be employed to significantly improve the computational efficiency and plan quality. We also propose a novel algorithm for the low-rank plus sparse matrix decomposition and demonstrate the effectiveness of the new algorithm on radiotherapy problems as well as background separation problems in computer vision.

3 - Optimal Dose Scheduling with Variable Fraction Sizes and Intervals via Sequential Mixed-Integer Convex Programming

Anqi Fu, Memorial Sloan Kettering, New York, NY, United States, Joseph O. Deasy, Zeno Gouw, Jeho Jeong

Hypofractionated radiotherapy has proven successful in local disease control, particularly for early-stage lung cancer. However, treatment is typically delivered via a constant dose schedule, which ignores the impact of cellular resource limits and reoxygenation on tumor response. We developed a method for constructing dose schedules with variable fraction sizes and intervals that exploits the difference in radiosensitivity between proliferating and resource-starved (hypoxic) cells. Our method formulates the dose scheduling problem as a finite-horizon optimal control problem, where the states are the proliferating/hypoxic cell counts and the controls are the dose fractions. The state transition dynamics are defined by a tumor response model of cellular resource competition and radiosensitivity. Since these dynamics are nonlinear, we designed an iterative optimization algorithm that solves a sequence of convex surrogates to arrive at an approximate solution to the problem. The resulting schedules consist of an initial large "primer shot" dose, followed by a 1-2 week break, and concluding with a short sequence of small doses. The primer shot targets the proliferating cells at the beginning since they are most radiosensitive; during the treatment break, these cells undergo post-mitotic cell death, freeing up resources so that hypoxic cells can reoxygenate. The newly reoxygenated cells have higher radiosensitivity and consequently require only 1-4 small doses to achieve the clinically-desired tumor control probability. Simulations on an early-stage lung cancer cohort showed that the primer shot schedule produces an average 6% reduction in local failures and 58% reduction in normal tissue BED3 compared to a constant dose schedule.

4 - Robust Optimization in Volumetric Modulated Arc Therapy for Prostate Cancer

Stoyan Hristov, University of Waterloo, Waterloo, ON, Canada, Houra Mahmoudzadeh, Ernest Osei, Johnson Darko

Volumetric modulated arc therapy (VMAT) is a type of radiation therapy in which a beam of radiation moves across an arc around the patient, continuously delivering radiation to a cancerous target. The goal of treatment plan optimization is to maximize radiation to cancerous cells and minimize damage to nearby healthy organs while ensuring that a set of mechanical constraints are met. VMAT is often used for prostate cancer treatment, which suffers from geometric uncertainty due to setup errors and changes in the patient's anatomy between treatments.

As such, we propose a VMAT model that uses robust optimization to optimize a minmax function and ensure that clinical criteria are met under uncertainty. The proposed model generates deliverable VMAT plans that are robust to a worst-case scenario. This talk will further motivate the research question, describe our methodology, and share preliminary results.

5 - Optimal Fractionation Plans in Temporally Feathered Radiation Therapy

Aysenur Karagoz, Rice University, Houston, TX, United States, Soheil Hemmati, Andrew Schaefer

Intensity-modulated radiation therapy (IMRT) remains the standard-of-care radiation therapy technique for most head-and-neck cancer (HNC) patients. Despite the relatively high survival rate of HNC (65-75%), the survivors often suffer from post-treatment long-lasting side effects, which negatively affect their quality of life. In an IMRT-based treatment plan, the daily doses delivered to the tumor and organs-at-risk (OARs) remain fixed based on the premise that the one-day inter-fraction suffices for each OAR's recovery. However, a novel treatment planning approach, referred to as temporally feathered radiation therapy (TFRT), proposes variable daily dose delivery to the OARs, aiming at increasing the recovery time of OARs at the cost of receiving higher doses once a week. While limited implementations of TFRT have shown promising results both in the simulations and in the clinic, they do not distinguish between the inherent radiobiological differences between the OARs for HNC patients, thus prescribing longer, yet uniform recovery time for all OARs. By recognizing variable radiosensitivity profiles across multiple OARs for HNC radiation therapy, we develop a mixed-integer nonlinear program (MINLP) that computes specific recovery time for each OAR to minimize the toxicity burden to the patients. Our proposed MINLP leverages the linear-quadratic model, which predicts each OAR's survival from radiation-induced damages. Our model is further flexible in addressing specific prioritization of OARs based on the clinician's preferences. The result of our model indicates a lower overall toxicity burden when OARs with higher radiosensitivity have longer recovery times under the TFRT framework.

TC62

Summit - Signature Room

Retail Price Optimization

Invited Session

The Practice Section of INFORMS

Chair: Elizabeth Cubbage, Petco, Fort Collins, CO, 80528, United States

1 - Using Consumer Price Perception to drive Optimal Pricing Strategy

Elizabeth Cubbage, Western Union, Fort Collins, CO, United States

Consumers often form price perceptions based on a limited set of goods rather than their entire basket. This research identifies key consumer price perception driving good types and the implications for pricing strategy.

2 - Revolutionizing Retail Pricing using GenAI

Sanjeewa Naranpanawe, OATL, Morrisville, NC, United States

In today's highly competitive retail landscape, pricing strategies can make or break a business. Traditional methods of pricing are often too rigid, slow, and fail to account for the dynamic nature of consumer demand and market trends.

Large Language Models such as GPT-4, are powerful tools for analyzing consumer sentiment, extracting insights from data, and generating dynamic pricing strategies. By processing textual data from social media, customer reviews, and market reports, LLMs can identify emerging trends and shifts in consumer preferences, providing retailers with a competitive edge.

Additionally, LLMs can facilitate personalized pricing by analyzing individual customer behavior and preferences, enabling retailers to tailor prices and promotions to maximize customer satisfaction and loyalty. These models can also be integrated into customer service chatbots and virtual assistants, providing real-time price recommendations and support.

3 - Lift Estimates and Schedule Optimization for Trade Promotion Planning

Maarten Oosten, Vistex, Inc., Cary, NC, United States

Trade promotions consist of the multi-level promotional activities of a manufacturer. The manufacturer sells their products via independent retailers to the consumers. They offer these retailers promotional programs to support their sales efforts. The manufacturers must incentivize the retailers to execute the promotions. After all, if a promotion is not attractive to the retailer, there won't be a promotion. Therefore, the manufacturer should model the behavior of the consumers as well as the behavior of the retailers. In this presentation, we discuss the challenges this poses in both the estimation of the promotion effects and the optimization of the promotion schedule.

4 - The effects of constraints on markdown price optimization

Nicholas Wegman, Zebra Technologies , Oak Park, IL, United States

For fashion and seasonal items, taking the right markdowns are essential to driving incremental revenue and minimizing excess inventory. It is well established that price optimization software can improve upon this business process by suggesting better markdown prices. However, the effectiveness of an optimized price is heavily influenced by the optimization constraints set by the user. In this talk we will discuss the importance of and methods for understanding how much optimality is lost by applying different constraints

TC63

Regency - 601

Generative AI and Creative Problem Solving

Invited Session

Information Systems

Chair: Leonard Boussioux, UW Foster, Harvard, Seattle, WA, 98122, United States

1 - Placeholder

Alex Burnap, Yale University, New Haven, CT, United States

2 - Generative AI and Creative Problem Solving

Miaomiao Zhang, Harvard Business School, Boston, MA, United States

The rapid advances in generative artificial intelligence (AI) open up attractive opportunities for creative problem-solving through human-guided AI partnerships. To explore this potential, we initiated a crowdsourcing challenge focused on sustainable, circular economy business ideas and assessed the novelty, value, and creativity of solutions created by both the human crowd and the collaborative efforts of one human and AI. The challenge attracted 125 global solvers from various industries and the human-prompted AI solutions were generated using strategic prompt engineering. We recruited 300 evaluators to judge a randomized selection of 13 out of 234 solutions, totaling 3,900 evaluator–solution pairs. Our findings demonstrate that the solutions generated through human-AI collaboration matched the creativity of those from the human solvers. Whereas the human-AI solutions provided more value, the human-only solutions were more innovative—both on average and for highly novel outcomes. Our study explores the potential for incorporating “AI-in-the-loop” into creative problem-solving, offering a scalable and cost-efficient method for enhancing the early phases of innovation. Our research paves the way for future exploration of how AI can be integrated into creative processes to foster more effective innovation.

3 - Generative AI and Evaluations of Early-Stage Innovations

Leonard Boussioux, University of Washington, Foster School of Business, Seattle, WA, United States, Charles Ayoubi, Ian Chen, Justin Ho, Jacqueline Lane, Camila Lin, Pei-Hsin Wang, Miaomiao Zhang

The rise of generative AI has transformed creative problem-solving, necessitating reevaluating idea evaluation processes to effectively assess and select from the resulting abundance of solutions. This study investigates how human-AI collaboration can enhance early-stage evaluations of innovative solutions, examining the interplay between objective criteria, which are quantifiable and measurable, and subjective criteria, which are based on personal opinion and intuition. Partnering with MIT Solve, a marketplace for social impact innovation, we conducted a field experiment involving 72 experts and 156 community screeners evaluating 48 real-world solutions for the 2024 Global Health Equity Challenge for a total of 3,002 screener-solution pairs. We utilized GPT-4 to provide recommendations and explanations for screening decisions. Our experiment compared a human-only control condition against two AI-assisted treatments: a black box AI providing recommendations without explanations and an explainable AI offering both recommendations and rationales for its decisions. Our findings reveal that screeners strategically use AI insights, validating AI’s recommendations when they agree and scrutinizing AI recommendations when they disagree. Screeners assisted by AI were roughly 9 percentage points more likely to fail a solution than the control group, primarily influenced by AI’s more stringent failure recommendations. Data from interviews and mouse tracking reveal that AI explanations for subjective criteria led screeners to doubt their own human judgment and possibly over-rely on AI’s explanations. This research suggests a possible framework for human-AI collaboration in creative evaluation, where AI complements human judgment on objective criteria while humans retain primary responsibility for subjective assessments supported by AI insights.

4 - LLM for Demand Forecasting

Dingli Ma, University of Washington, Seattle, WA, United States

Prior literature suggests that human forecasters typically overreact to forecast errors in relatively stable environments but underreact to errors in relatively unstable environments. We train an LLM for demand forecasting and investigate whether the trained LLM can overcome systematic judgment biases.

TC64

Regency - 602

Shaping Perceptions and Identities in Social Media

Invited Session

Social Media Analytics

Chair: Reihane Boghrati, Arizona State University, Tempe, AZ, United States

1 - Intersectional stereotypes in everyday language: Evidence across race, gender, and social class

Tessa Charlesworth, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Kshitish Ghate, Aylin Caliskan, Mahzarin Banaji

Social group-based identities intersect. The meaning of “woman” is modulated by adding social class as in “rich woman” or “poor woman”. How does such intersectionality operate at-scale in everyday language of the internet and social media? Which intersections dominate (are most frequent)? What qualities (positivity, competence, warmth) are ascribed to each intersection? Here, we make it possible to address such questions by developing a new stepwise procedure, *Flexible Intersectional Stereotype Extraction* (*FISE*), applied to word embeddings (*GloVe*; *BERT*) trained on billions of words of English Internet text revealing insights into intersectional stereotypes. First, applying FISE to occupation stereotypes across intersections of gender, race, and class showed alignment with ground-truth data on occupation demographics, providing initial validation. Second, applying FISE to trait adjectives showed strong androcentrism (*Men*) and ethnocentrism (*White*) in dominating everyday English language (e.g., *White+Men* are associated with 59% of traits; *Black+Women* with 5%). Associated traits also revealed intersectional differences: advantaged intersectional groups, especially intersections involving *Rich,* had more common, positive, warm, competent, and dominant trait associates. Together, the empirical insights from *FISE* illustrate its utility for transparently and efficiently quantifying intersectional stereotypes in existing large text corpora, with potential to expand intersectionality research across unprecedented time and place. This project further sets up the infrastructure necessary to pursue new research on the emergent properties of intersectional identities.

2 - Does Authenticity Matter in the Age of AI Music?

Yifei Wang, University of New Hampshire, Durham, NH, United States, Yunfei Wang, Jui Ramaprasad

<p break-words""="">This study examines how the release of the generative AI model GPT-4 in March 2023 impacted music consumption patterns, focusing on the role of perceived authenticity as a key driver. The advent of AI music composition has sparked debates around copyright, originality, and AI's potential threat to artistic integrity. We argue that GPT-4's release motivated consumers to increasingly seek out music perceived as more authentic through raw production and earnest self-expression. Analyzing a novel dataset from Spotify's influential "New Music Friday" playlist, we investigate demand shifts at both the genre and artist levels using difference-in-differences models with propensity score matching. We find GPT-4's release increased weekly streaming by 5.19% for the more authentic Electronic genre compared to Pop. At the artist level, original artists (credited as composers/producers) experienced a 3.94% additional streaming increase over non-original artists after GPT-4's launch. We further explore the interaction of originality and proficiency (measured by major label association) as composite authenticity indicators. Our findings provide early evidence on how consumers assess music authenticity in the AI era, with authentic genres and original, lower-proficiency artists receiving amplified demand following GPT-4's release. This offers insights for industry stakeholders navigating the changing landscape as generative AI shapes evolving authenticity perceptions.

3 - A Multi-Modal Assessment of User Engagement in How-to Videos on YouTube

Maryam Rahmani Moghaddam, University of Delaware, Newark, DE, United States, Xiao Liu, Anjana Susarla, Rema Padman

YouTube how-to videos have proven to be an effective means of patient education. However, producing engaging and high-quality how-to videos is not easy. This study examined YouTube videos related to diabetes-related topics to provide practical insights into how to create high-quality videos that facilitate patient understanding and engagement. Drawing on the literature on cognitive load theory and the cognitive theory of multimedia learning we propose a set of visual and verbal features that can affect the cognitive load of learners while learning new skills or knowledge. In particular, we construct features that capture visual complexity in terms of color and objects displayed. For narratives, we focus on the amount and flow of information conveyed. In addition, eleven acoustic features such as pitch and intensity are used to quantify vocal modality. Our study indicates that vocal features have a greater impact on user engagement as measured by the number of views and likes. Findings from the study can benefit health consumers, patient educators, and health organizations.

4 - Impact of Weight Stigmatization on Avatar Selection in the Metaverse

Sungeun Han, Arizona State University, Tempe, AZ, United States, Reihane Boghrati

While weight discrimination manifests in diverse dimensions including workplaces, and schools, people have been facing weight bias online with the rise of social media. The increased anonymity makes it easy to post memes, and comments that would perpetuate body shaming and weight prejudice from behind the screen. It is not limited to social media, in more real-time interactions virtual realm including metaverse, distorted weight-related attitudes and beliefs can be more problematic. In the metaverse where users have complete autonomy to create their own digital avatars, body shaming, and weight stigma can influence the way users represent themselves in the digital realm. Grounded in self-discrepancy theory, previous studies found that people tend to reconstruct their digital identities to fulfill their ideal selves. Through the lens of self-discrepancy theory, this study delves into the influence of weight stigmatization on consumer behavior in avatar selection within the metaverse. Furthermore, we investigate how weight-stigmatized users cope with stigmatization through compensatory consumption, such as purchasing luxury items in the virtual realm as a way of coping with stigmatization. Additionally, we seek to understand how avatar selection behavior may vary depending on different social contexts, where users interact with friends or strangers. As the metaverse expands, there is a rising concern over the lack of diverse avatar options, with a noticeable predominance of thin avatars. Through our study, we aim to promote a more inclusive digital environment emphasizing the importance of avatar diversity by offering an in-depth understanding of how digital identity intertwines with weight stigma.

5 - Unveiling the Heart of a Social Movement: Analyzing the Drivers of User Engagement on Social Media for #Blacklivesmatter

Adrija Majumdar, Indian Institute of Management Ahmedabad, Ahmedabad, India, Chellapilla Deep Prakash, Rajiv Gopinath

Social movements fundamentally reshape societal norms through their advocacy for justice, equality, and human rights. Prior research has explored social media's role in user engagement, emphasizing cultural influences on content creation and its subsequent effects on communication, information spread, and societal change. This study builds on existing literature by specifically focusing on how different content types—sensationalist, justice-oriented, objective, or action-oriented—influence user engagement within the context of social movements. Employing the Uses and Gratifications Theory, which posits that individuals selectively use media to satisfy distinct needs such as information, entertainment, and social integration, our analysis centers on social movement-related posts tagged with #BlackLivesMatter during the heightened periods of activism in August 2017 and May 2020. Our methodological framework includes exploratory data analysis, the creation of a specialized dictionary, feature engineering, and regression analysis to identify key engagement drivers. The results indicate that sensationalist content markedly increases user engagement, with additional factors like follower count, tweet length, and multimedia presence also significantly contributing. This research not only deepens insights into the dynamics of media consumption during pivotal social movements but also offers guidelines for crafting effective digital content. Highlighting the importance of analyzing engagement drivers, this study shows that increased engagement fosters greater awareness and more efficient information dissemination, thereby amplifying the influence of social movements. Our findings contribute to the ongoing discourse on the pivotal role of social media in supporting social movements.

TC65

Regency - 603

To Accredit or Not To Accredit: That is the Question

Panel Session

Education Outreach

Co-Chair: Ken Murphy, Merage School - UC Irvine, 4293 Pereira Drive, Irvine, CA, 92697, United States

1 - Moderator Panelist

Ken Murphy, Merage School of Business, UC Irvine, Irvine, CA, United States

2 - Panelist

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

This panel will discuss academic program accreditation for business analytics/data science, ABET, and standards alignment with the Certified Analytics Professional (CAP). Panelists will share their experience with advocating for the value of accreditation, building appropriate administrative support, and executing certification acquisition and maintenance activities. Opportunities for INFORMS to grow with respect to academic program accreditation will also be considered.

3 - Panelist

Alper Atamturk, UC Berkeley, Berkeley, CA, United States

4 - Panelist

Bill Griffin, INFORMS, Catonsville, MD, United States

5 - Panelist

Naser Nikandish, Johns Hopkins Carey Business School, Washington, DC, United States

TC66

Regency - 604

Machine Learning in Decision-Making: Strategies and Interpretability

Invited Session

Artificial Intelligence

Chair: Wangsheng Zhu, Hong Kong University of Science and Technology , Hong Kong, N/A, Hong Kong

1 - Care for the Mind Amid Chronic Diseases: An Interpretable AI Approach Using IoT

Jiaheng Xie, University of Delaware, Newark, DE, United States

Health sensing for chronic disease management creates immense benefits for social welfare. Existing health sensing studies primarily focus on the prediction of physical chronic diseases. Depression, a widespread complication of chronic diseases, is however understudied. We draw on the medical literature to support depression prediction using motion sensor data. To connect human expertise in the decision-making, safeguard trust for this high-stake prediction, and ensure algorithm transparency, we develop an interpretable deep learning model: Temporal Prototype Network (TempPNet). TempPNet is built upon the emergent prototype learning models. To accommodate the temporal characteristic of sensor data and the progressive property of depression, TempPNet differs from existing prototype learning models in its capability of capturing the temporal progression of depression. Extensive empirical analyses using real-world motion sensor data show that TempPNet outperforms state-of-the-art benchmarks in depression prediction. Moreover, TempPNet interprets its predictions by visualizing the temporal progression of depression and its corresponding symptoms detected from sensor data. We further conduct a user study to demonstrate its superiority over the benchmarks in interpretability. This study offers an algorithmic solution for impactful social good - collaborative care of chronic diseases and depression in health sensing. Methodologically, it contributes to extant literature with a novel interpretable deep learning model for depression prediction from sensor data. Patients, doctors, and caregivers can deploy our model on mobile devices to monitor patients' depression risks in real-time. Our model's interpretability also allows human experts to participate in the decision-making by reviewing the interpretation of prediction outcomes and making informed interventions.

2 - Interpretable Feature Selection for Predictive Analytics: A Bayesian Perspective

Wei Feng, University of Maryland, College Park, MD, United States, Kunpeng Zhang

Feature selection aims to select a pertinent subset of features, thereby shrinking dimensionality of dataset to enhance machine learning model efficacy. It has seen widespread success across various domains. A major hurdle, however, is the nature of ‘black-box’ in most prior studies: the importance and interplay of chosen features are often nebulous, posing significant risks in critical areas like finance and healthcare. To tackle this, several well-known model-agnostic interpretative frameworks like SHAPLEY and LIME have been suggested to elucidate the relationships between features and outcomes. This is not satisfactory, due to the fact that two other key research gaps remain unfilled: automatically determining the optimal size of selected features and unraveling the inter-feature relationships. This paper introduces a novel Bayesian-based feature selection framework that integrates structural learning with the Markov blanket theory in predictive models. This method not only selects informative features but also shows the pathways leading to outcomes, therefore boosting interpretability. The proposed framework was empirically validated against established benchmark models using both real and simulated data. Results showed that (i) using the selected features from our method achieves better performance than several state-of-the-art baselines, and even comparable to using full features. (ii) the statistical 'causal' relations among features are also generated, which is very useful in practice. (iii) On simulated data, the framework accurately identified predefined relationships among variables, achieving over 80% accuracy with few false positives.

3 - Investigating ideational shifts across technological generations in the telecom industry

Tzu-Wen Lin, The University of Texas at Arlington, Arlington, TX, United States, Sridhar Nerur, Jayarajan Samuel, Ruochen Liao

The transition from 3G to 5G technologies in the telecommunications industry has sparked a debate on whether it represents a fundamental paradigm shift or merely an incremental technological advance. This research delves into this transformation, utilizing patent data and advanced techniques from network science and text analytics to elucidate the nature of ideational changes in the innovative landscape of telecommunications.

Employing methods such as machine learning prediction and topic modeling, the study evaluates lexical differences across telecommunication generations, thereby offering insights into the industry's level of disruption. Furthermore, this research investigates the pivotal role of knowledge as a source and the nature of knowledge interchange between firms, tracing the innovative evolution of top-performing companies from the 3G era to the present time. By unraveling the intellectual structure of innovations in the telecommunications domain, the research provides invaluable perspectives on the dynamics of technological change and the significance of knowledge exchange in shaping the industry. Through this comprehensive endeavor, the study aims to deepen our understanding of the cross-generational shifts occurring within the telecommunications industry, shedding light on the nature and implications of the transition from 3G to 5G.

4 - Navigating the Unknown: Optimizing Advertising Strategies in the Face of Ad Auction Uncertainty

Wangsheng Zhu, Hong Kong University of Science and Technology , Hong Kong, Hong Kong, Shaojie Tang, Vijay Mookerjee

Online advertising has emerged as a popular method for firms to promote their products. Unlike the offline context, where advertising slots are predetermined, most online advertising slots are sold through real-time auctions. The outcome of these auctions is influenced by various factors, including the number of competing advertisers. Consequently, firms launching advertising campaigns face uncertainty regarding the number of slots they will secure.

While previous studies have examined optimal advertising strategies assuming accurate knowledge of ad exposure, the impact of slot acquisition uncertainty during ad auctions remains unexplored. Building upon existing literature, we propose a diffusion equation that explicitly models the randomness associated with slot acquisition in ad auctions. By validating our model using a real-world dataset, we demonstrate that firms currently operate sub-optimally by overlooking slot acquisition randomness when launching ad campaigns. Furthermore, we illustrate how our model can be utilized to derive optimal advertising strategies that account for slot acquisition uncertainty, leading to improved profitability for advertisers. Additionally, apart from maximizing expected sales, firms may also aim to maintain low sales variance for effective inventory management. Our model accommodates this consideration, allowing firms to derive optimal advertising strategies that strike a balance between mean and variance of sales.

5 - From Rankings to Decisions: Deep Learning-Based Aggregation Methods

Shuzhang Cai, The University of Texas at Dallas, Richardson, TX, United States, Shaojie Tang

Many decision-making processes, such as participatory budgeting or candidate recruitment, involve collecting and aggregating participants' voting results. Obtaining accurate utility scores from voters can be challenging; instead, they often use rankings to indicate their preferences. Different methods of aggregating these rankings can lead to varying outcomes. In this paper, we aim to develop a deep learning-based system to effectively aggregate these rankings and produce a consensus decision. Our approach addresses the practical difficulties of utility estimation while enhancing the reliability of the final decision.

TC67

Regency - 605

Personalized and Fair AI: Advances in Federated Learning and LLMs

Contributed Session

Chair: Chong Oh, Eastern Michigan University, Ypsilanti, MI, United States

1 - Tailoring Large Language Model to Satisfy Fine-Grained User Needs: A Novel Deep Learning Method

Shujie Liu, Zhejiang University, Hangzhou, China, People's Republic of, Gang Chen, Xiangwei Kong

The powerful language comprehension and content generation capabilities of large language models (LLMs) imply great potential for commercial applications. However, when applied to domain-specific conversation scenarios, they often yield vague and generic responses, failing to fulfill personalized user needs. In this paper, we frame personalized user needs as need coverage and need specificity based on information processing theory, and strive to tailor LLM to satisfy fine-grained user needs. To achieve this, we propose a novel deep learning method named MSA-PAT, which synergizes two components, i.e., MSA for inferring multi-granularity user needs based on conversation text and PAT for tailoring the LLM's generated responses toward fine-grained user needs satisfaction through prompt adaptation and self-supervised LLM fine-tuning. Empirical evaluation in two question-answer case studies demonstrates the satisfactory response generation capability of MSA-PAT toward fine-grained user needs. Ablation, sensitivity analyses and case studies provide insights into how MSA-PAT generates responses to satisfy fine-grained user needs.

2 - Personalized Federated Learning via Learning Shared Representation: Statistically Optimal Rates

Xiaochun Niu, Northwestern University, Evanston, IL, United States, Lili Su, Jiaming Xu, Pengkun Yang

Personalized federated learning enables clients and a central server to collaboratively learn a shared model while allowing individual clients to fine-tune and adapt the model to their specific needs privately. Current literature primarily focuses on algorithmic advancements and assumes that the local data volume far exceeds the model dimension, leaving the optimal sample complexity unresolved.

In this paper, we determine the optimal sample complexity when clients share a common low-dimensional linear representation. We characterize statistically optimal rates for both unsupervised and supervised learning settings. Specifically, we establish information-theoretic lower bounds on estimation errors and design a federated method-of-moments estimator with local averaging to achieve these bounds. Our results demonstrate that the optimal error rates match those in centralized learning with a moderate number of agents and local dataset sizes. However, a statistical penalty arises due to distributed computing when there are many agents with insufficiently large or heterogeneous local datasets.

3 - Enhancing Group Fairness in Federated Learning Through Personalization

Yifan Yang, The Ohio State University, Columbus, OH, United States, Ali Payani, Parinaz Naghizadeh

Instead of producing a single global model for all participating clients, personalized Federated Learning (FL) algorithms aim to collaboratively train customized models for each client, enhancing their local accuracy. For example, clients could be clustered into different groups in which their models are similar, or clients could tune the global model locally to achieve better local accuracy. In this paper, we investigate the impact of personalization techniques in the FL paradigm on local (group) fairness of the learned models, and show that personalization techniques can also lead to improved fairness. We establish this effect through numerical experiments comparing two types of personalized FL algorithms against the baseline FedAvg algorithm and a baseline fair FL algorithm, elaborate on the reasons behind improved fairness using personalized FL methods, and provide analytical support under certain conditions. Based on these findings, we further propose a fairness-aware clustered FL algorithm and analyze the convergence rate of this algorithm in a convex and smooth setting. Our experiment results demonstrate that our proposed algorithm has improved fairness compared to classical clustered FL, which could lead to an improved fairness-accuracy tradeoff.

4 - Strategic Retraining of LLMs: Variance, Value, and Competitive Dynamics

Ehsan Valavi, Massachusetts Institute of Technology, Cambridge, MA, United States, Joel Hestness, Marco Iansiti, Newsha Ardalani, Feng Zhu, Karim Lakhani

This paper investigates the strategic and operational implications of retraining large language models (LLMs) in diverse business contexts. By analyzing user-generated content from Reddit, we identify significant variations in data perishability and retraining frequencies across different domains. Our findings indicate that rapidly evolving topics, such as politics and world news, require more frequent updates, while stable areas like relationship advice and history benefit from less frequent retraining. We introduce the concept of "data effectiveness" to measure predictive power over time, emphasizing the necessity for continuous access to fresh data in dynamic environments.

Our study highlights the complementary role of human expertise in providing up-to-date data, especially in fast-changing contexts where LLM performance relies heavily on current information. In more stable domains, historical data retains its value, allowing for the use of more sophisticated models. We discuss the implications of data perishability for AI strategies, advocating for flexible data management systems that support frequent model updates to maintain relevance.

This research contributes to the literature on AI economics, data valuation, and strategic management, offering insights into optimizing LLM retraining to enhance business competitiveness and productivity. It underscores the importance of adapting LLM deployment strategies to the pace of change within specific domains, ultimately shaping the future of data-driven decision-making and competitive advantage.

5 - Data sharing through federated learning in supply chain: a game theoretical analysis

Yonghua Ji, University of Alberta, Edmonton, AB, Canada, Huimin Ma

Federated learning has been used to improve operations and protect privacy in various industries. This study focuses on the use of federated learning in the supply chain context. By utilizing local models provided by its clients, an equipment manufacturer could use federated learning technology to train a global model and improve the operations of its equipment in its clients' factories. Through a game theoretical model, we study how the equipment manufacturer can motivate the clients to participate in its federated learning platform. Furthermore, we study the effect of federated learning on the clients' and the manufacturer's profits and social welfare.

6 - Enhancing Mental Health Support Through AI: A Study on High-Contingency Chatbots and Social Support Theory

Chong Oh, University of Utah, Salt Lake City, UT, United States, Matthew Pecsok, Carina Pals, Rishi Ramesh

This study aims to explore the efficacy of AI chatbots in providing social support to college students facing academic stress, framed within the context of Social Support Theory. We will conduct an experiment comparing high-contingency and low-contingency chatbots with human support providers. Participants will be randomly assigned to interact with one of these support types and subsequently complete surveys evaluating perceived support, cognitive reappraisal, emotional validation, and emotional distress. High-contingency chatbots, leveraging advanced natural language processing (NLP) to deliver context-aware and personalized responses, will be contrasted with low-contingency chatbots that offer generic, scripted replies. Our hypothesis is that high-contingency chatbots will be perceived as more competent and warmer, thus more effective in reducing emotional distress through enhanced cognitive and emotional engagement. This research seeks to contribute to Information Systems (IS) by demonstrating how AI-driven, adaptive communication technologies can be designed to emulate human-like support and improve mental health outcomes. By integrating Social Support Theory with cutting-edge AI, this study aims to provide actionable insights for developing intelligent, empathetic chatbot systems in the mental health domain.

TC68

Regency - 606

Mitigating Bias and Enhancing AI Collaboration

Contributed Session

Chair: Aida Khosh Raftar Nouri, Memorial university, 200 Crestview Church Road, Apartment #1405, Warner Robins, GA, 31088, United States

1 - To Automate or Not? How Open Collaboration Communities Decide on Bot Adoption

Lei Nico Zheng, University of Massachusetts Boston, Boston, MA, United States, Zihan Chen, Feng Mai

AI tools are widely used in open collaboration communities nowadays, yet factors influencing their adoption are not well understood. This study investigates the factors influencing bot adoption on English Wikipedia. Leveraging the advanced capabilities of Large Language Models (LLMs) on NLP tasks, we systematically extract human-computer interaction (HCI) factors from 2,155 Wikipedia Bot Request for Approval pages. We identify ten themes emerged form community discussion concern three main aspects: Technological Factors, User Interaction and Experience, and Platform Governance. We further assess the comparative importance of each theme in the community's bot adoption decisions. Our findings reveal that Technological Factors play the most significant role, followed by User Interaction and Experience and Platform Governance Standards. This study contributes quantitative evidence to understanding bot adoption in open collaboration communities and proposes a novel, generalizable LLM-based approach for extracting and summarizing themes from large text corpora.

2 - Advancing Decision Support with Generative AI and Human-AI Collaborations in Multi-Agent Systems

Seonho Woo, Purdue University, West Lafayette, IN, United States, Young-Jun Son, Seung ho woo

This study proposes a novel decision-making framework leveraging generative AI and multi-agent systems to enhance human-AI collaborations. This framework employs synthetic data to train AI agents and circumvents privacy concerns while magnifying the robustness of experiments in simulated experiments. Incorporating a generative AI enables the system to create a variety of scenarios dynamically, mimicking complex contexts in real-world decision-making and allowing for extensive testing and refinement of AI behaviors in controlled, safe settings.

This research focuses on developing decision support in the multi-agent system in which AI agents generate and evaluate potential decisions so that they collaboratively work with human inputs to optimize outcomes. This setting improves decision accuracy and enhances the AI-system adaptability simultaneously to unpredictability and human strategies. Various complexity levels in decision-making tasks are included in experiment designs and are utilized to assess the effectiveness of human-AI collaboration and cooperation of AI agents and human operators in real time.

Our goal is to signify the effect of a generative AI-enhanced multi-agent framework for decision support that can enhance the decision-making process for the system where human plays a significant role and AI agents are advantageous concurrently. This approach enables human traits to be benefited from AI capabilities in human-run operating systems, and this leads to more efficient and informed decision-making processes in various applications with higher performance.

3 - Mitigating Omitted Variable Bias through the Design of Data Collection Interfaces

Aida Nouri, Memorial university , Warner Robins, GA, United States

This study explores the potential impact of omitted variable bias in machine learning (ML), which may arise from traditional class-based data collection methods. Such methods, by limiting data collection to predefined categories, might inadvertently overlook important features of entities, leading to possible biases. We investigate this phenomenon by contrasting class-based data collection with a flexible instance-based data collection approach, hypothesizing that the latter can capture a broader spectrum of attributes and more accurately reflect real-world phenomena.

Using a comparative experimental design, participants will interact with objects in videos, categorizing and describing them through both class-based and instance-based methodologies. The class-based method may confine data to predefined attributes, potentially limiting the richness and variability of the collected data. In contrast, the instance-based method is expected to allow detailed, unrestricted descriptions, leading to a richer dataset that better captures the nuances and complexities of real-world entities.

Our study aims to employ sensitivity analysis to investigate the extent of omitted variable bias under each data collection methodology, examining the robustness of model estimations with variable modifications. By doing so, we seek to understand how different data collection practices influence the quality and reliability of ML models. This research contributes to the broader understanding of data collection practices in ML and other data-driven fields, advocating for more inclusive and dynamic methods for data collection. By adopting more flexible and comprehensive data collection approaches, we aim to improve the overall reliability and validity of ML models and other analytical tools.

TC69

Regency - 607

SpORts III

Invited Session

spORts

Co-Chair: Keith Willoughby, University of Saskatchewan, Saskatoon, SK, Canada

1 - Analysis of the Push Phase of the 2022 Olympic Bobsled Competition

Aaron Hoskins, California State University, Fresno, Fresno, CA, United States

The push phase is critical in bobsled competition because extra speed at the top of the mountain results in a significant reduction in the time required to finish the run. This work analyzes the dynamics of all teams from the 2022 Olympic Bobsled competition to determine characteristics that separated the teams from each other.

2 - Post COVID, Higher-Rated Women's Teams Won More Often and Higher-Rated Men's Teams Won Less Often, Likely Due to Women's Greater Team Cohesiveness

Raymond Stefani, California State University, Long Beach, Lake Forest, CA, United States

A pre-COVID 2007-2019 data base was gathered to study all team sports that had international recognition, had an official rating system published by the governing federation and had a world championship (WC). The higher-rated women's teams won only 0.25% more than the men's teams. Post-COVID, 11 WCs were contested during 2021and 2022, and early in 2023, in the five sports using the same rating system as pre-COVID and contested by both men and women. In all five sports, the percentage of games won by the higher-rated men's team was lower post-COVID: curling (-6.1%), rugby 7s (-12.9%), T20 Cricket (-8.9%), basketball (-12.6%), and rugby union (-2.0%). However, in all five sports, the women's higher-rated teams had higher percentages post-COVID: curling (+7.3%), rugby 7s (+6.2%), T20 cricket (+7.0%), basketball (+5.0%) and rugby union (+0.4%). A review of the coaching literature suggests that during the COVID-era restrictions, adverse conditions triggered an increase in women's social cohesiveness, which drove increased team cohesiveness, creating increased cooperative learning and execution of tactical skills, whereas men emerged from those adverse training conditions with less team cohesiveness and predictability.

3 - The Canadian men's curling championship: Simulating the path to the podium

Keith Willoughby, University of Saskatchewan, Saskatoon, SK, Canada, Kent Kostuk

For nearly a century, Canadian curlers have annually competed in a national championship known as the "Brier". Regulations dictate that the championship entrants include at least one team from each Canadian province and territory. Starting in 2014, curling officials gave an automatic "Brier" bye to the previous year's winner. The championship field was recently increased to an 18-team event with the addition of "wild card" entrants. We develop an analytical model to determine the probability that any team captures the national championship trophy. In addition, we investigate the fairness of automatic byes for the previous year's champion.

4 - Predicting March Madness Upsets with Gradient Boosting Models

Gabriel Forthofer, USAFA, Colorado Springs, CO, United States

The NCAA Men's Basketball Tournament, known as March Madness, is a high-stakes, single-elimination college basketball event that captivates audiences with its unpredictable upsets and "Cinderella stories." We apply the machine learning technique of Gradient Boosting Model (GBM) to predict these upsets. Utilizing a dataset spanning from 2002 to 2022, we compile team statistics and match outcomes to train our model. We evaluate model performance using accuracy and feature importance metrics, revealing insights into the dynamics of March Madness upsets. Preliminary results indicate a promising capability of the GBM in predicting upsets with significant accuracy, offering a novel approach to understanding and forecasting outcomes in one of America's most thrilling sports events. This study not only contributes to the field of sports analytics but also enhances the understanding of applying advanced predictive models in team strategy development.

TC70

Regency - 701

Stochastic Resources Bidding in Flexibility Markets

Invited Session

ENRE: Electricity

Chair: Jalal Kazempour, Technical University of Denmark, Kgs. Lyngby, Denmark

Co-Chair: Lesia Mitridati, Technical University of Denmark, Kgs Lyngby, Denmark

1 - A Risk-Aware Approach to Optimizing Frequency Reserve Dimensioning

Alessandro Zocca, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

One of the main responsibilities of Transmission System Operators (TSOs) that operate an electric grid, is to maintain a certain frequency (e.g., 50 Hz in Europe). To be able to do so, TSOs have created several products that are called "frequency supporting ancillary services". The Frequency Containment Reserve (FCR) is one of these ancillary service products and to some TSOs it is unclear how to dimension it. In our work, we study the risk effect of changing the often fixed FCR obligation to a dynamic approach. We take the Nordic synchronous area in Europe as case study and use a diffusion model to capture its frequency development and introduce a controlled mean reversal parameter to assess changes in FCR obligations, in particular for the FCR-D ancillary service product. We establish closed form expressions for exceedance probabilities and use real life (frequency) data to study the risk effects numerically.

2 - Reach and Hold Flexibility Characterization for Aggregations of Thermostatically Controlled Loads

Mazen Elsaadany, University of Vermont, Burlington, VT, United States

Thermostatically controlled loads (TCLs) have the potential to be flexible and responsive loads to be used in demand response (DR) schemes. With increasing renewable penetration, DR is playing an increasingly important role in enhancing power grid reliability. The aggregate demand of a population of TCL's can be modulated by changing their temperature thermostat setpoint. When and/or what proportion of the population sees the thermostat setpoint change determines the change in aggregate demand. However, since the TCL population is finite, not all changes in aggregate demand can be maintained for arbitrarily long periods of time. The dynamic behavior of a TCL fleet is modeled and *used to characterize the set possible changes in aggregate demand that can be reached and the corresponding time for which the demand change can be held, for a given change in thermostat setpoint. This set is referred to (in this work) as the 'reach and hold set' of a TCL fleet. Furthermore, the effect of the size of the thermostat setpoint change on the reach and hold set is analyzed. The characterized set is then validated through simulation using both the population TCL models and individual TCL micro-models.*

3 - Chance-constrained bidding strategy for a flexibility aggregator of stochastic distributed units in Nordic FCR-D markets

Torine Reed Herstad, Technical University of Denmark, Kgs. Lyngby, Denmark, Jalal Kazempour, Bert Zwart, Lesia Mitridati

The provision of ancillary services can offer an additional source of revenues for flexible loads, including those with stochastic power consumption baselines, such as Electric Vehicles (EVs). In order to harness this flexibility, the Danish transmission system operator, Energinet, sets a minimum reliability requirement of 90% for stochastic flexibility providers participating in the Nordic primary ancillary service (FCR-D) market. In this context, we study the optimal bidding strategy of an aggregation of EVs in the FCR-D market. The minimum reliability conditions are modelled using joint chance-constraints, and novel exact solution methods using extreme value theory are proposed. The dependence between the upward and downward flexibility at different hours is investigated using copulas and it is discussed to which extent the assumption of independence leads to over-bidding. The performance of the proposed solution methods is analyzed empirically and compared to a Sample Average Approximate (SAA) solution method on a realistic case study consisting of aggregated charging boxes in Danish households.

4 - Market Operations & Battery Storage Conundrums

Ben Hobbs, Johns Hopkins University, Baltimore, MD, United States

The rapid penetration of storage batteries in bulk power markets pose challenges to the traditional thermal plant paradigms of optimal dispatch and market power mitigation. Present market designs are not up to these challenges for several reasons, including stochasticity, abbreviated time horizons, the need to estimate opportunity costs, presence of price caps, and degradation costs. These challenges and possible responses to them by market designers are reviewed.

TC71

Regency - 702

Applications of Machine Learning and AI for Physical and Mental Health

Invited Session

Data Mining

Chair: Zhe Gao, Shanghai Normal University, Shanghai, N/A, China, People's Republic of

Co-Chair: Jiajing Huang, Arizona State University, Reseda, 85281

1 - Healthy Bio-Core: An Instance Selection Framework for Enhancing Biomedical Temporal Data Classification

Abhidnya Patharkar, ARIZONA STATE UNIVERSITY, Tempe, AZ, United States

Numerous existing machine learning methods engage in instance selection, specifically instance filtering, to eliminate outliers or samples that are on the edge. We present a new technique for conducting instance selection on multivariate time series data. We explore the implementation of this technique in identifying a representative subset of the healthy class in three biomedical temporal data sets. This selected subset is referred to as a 'Healthy Bio-Core' (HBC). Our approach involves the utilization of dynamic time warping and Gaussian distribution to create the HBC. Unlike conventional methods that often rely on clustering algorithms such as the k-nearest neighbors algorithm for instance filtering, which necessitates specifying the parameter k, our algorithm eliminates the need for parameter specification by leveraging the classification accuracy of the ROCKET (Random Convolutional Kernel Transform) algorithm. The results demonstrate enhanced performance when using HBC compared to all healthy data. To further validate the performance, we employ three classifiers: HIVE-COTE (Hierarchical Vote Collective of Transformation-based Ensembles), MUSE (Multi- variate Unsupervised Symbols and Derivatives) and DTW-NN (Dynamic Time Warping with Nearest Neighbor), assessing accuracy and F-1 score of classification for HBC versus abnormal data, and healthy versus abnormal data.

2 - A Multi-Scale Self-Corrected Masked Autoencoder with Brain Tumor Applications

Hairong Wang, Georgia Institute of Technology, Atlanta, GA, United States

Glioblastoma (GBM) is one of the most aggressive and lethal human cancers. Intra-tumoral genetic heterogeneity poses a significant challenge for treatment. Biopsy is invasive, which motivates the development of non-invasive, MRI-based machine learning (ML) models to quantify intra-tumoral genetic heterogeneity for each patient. This capability holds great promise for enabling better therapeutic selection to improve patient outcome.

3 - Blobcut: a Contrastive Learning Method to Support Small Blob Detection in Medical Imaging

Teng Li, Arizona State University, Tempe, AZ, United States

Medical imaging-based biomarkers derived from small objects (e.g., cell nuclei) play a crucial role in medical applications. However, detecting and segmenting small objects (a.k.a. blobs) remains a challenging task. In this research, we propose a novel 3D small blob detector called BlobCUT. BlobCUT is an unpaired image-to-image (I2I) translation model that falls under the Contrastive Unpaired Translation paradigm. It employs a blob synthesis module to generate synthetic 3D blobs with corresponding masks. This is incorporated into the iterative model training as the ground truth. The I2I translation process is designed with two constraints: (1) a convexity consistency constraint that relies on Hessian analysis to preserve the geometric properties and (2) an intensity distribution consistency constraint based on Kullback-Leibler divergence to preserve the intensity distribution of blobs. BlobCUT learns the inherent noise distribution from the target noisy blob images and performs image translation from the noisy domain to the clean domain, effectively functioning as a denoising process to support blob identification. To validate the performance of BlobCUT, we evaluate it on a 3D simulated dataset of blobs and a 3D MRI dataset of mouse kidneys. We conduct a comparative analysis involving six state-of-the-art methods. Our findings reveal that BlobCUT exhibits superior performance and training efficiency, utilizing only 56.6% of the training time required by the state-of-the-art BlobDetGAN. This underscores the effectiveness of BlobCUT in accurately segmenting small blobs while achieving notable gains in training efficiency.

4 - Interpretable Deep Learning Framework for Understanding Molecular Changes In Human Brains with Alzheimer&Rsquo;S Disease

Maitry Ronakbhai Trivedi, Arizona State University, Tempe, AZ, United States

The potential of Artificial Intelligence (AI) in investigating the genetic irregularities spanning various brain regions affected by Alzheimer's disease (AD) is yet to be fully explored. Specifically, there exists a significant gap in identifying common and specific transcriptomic signatures that characterize AD-related cellular and molecular processes. This undertaking shows potential in illuminating new insights into the biology of the disease, aiding in the identification of biomarkers, and facilitating the discovery of targets. Therefore, we propose an extensive deep-learning approach that consists of Multi-Layer Perceptron (MLP) to classify the AD vs controls using the bulk-RNA sequence data from three different brain regions in the Religious Orders Study/Memory and Aging Project (ROSMAP) from the RNAseq Harmonization Study of the AMP-AD consortium. Utilizing predicted embeddings of the MLP, we model the distribution of expression profiles as a progressive disease trajectory across distinct brain regions. To enhance the interpretability of MLP models, we employ SHapley Additive exPlanations (SHAP) values and identify the most significantly AD-implicated genes, which serve as key nodes for subsequent gene co-expression network analysis. Our models exhibit robust performance in classification and prediction, validated across two external datasets: the Mayo RNA-seq (MAYO) cohort and the Mount Sinai Brain Bank (MSBB) cohort of AMP-AD. This research emphasizes the transformative capacity of AI in unraveling the complex molecular mechanisms of AD, surpassing conventional approaches to reveal novel insights into the disease's pathogenesis.

TC72

Regency - 703

High-Dimensional Data Analytics for Complex Systems

Invited Session

Data Mining

Chair: Anyi Li, Auburn University, Auburn, 36849, United States

Co-Chair: Jia Liu, Auburn University, Auburn, United States

1 - Deep Learning for Overheating Prediction in 3D Printing: A CNN-LSTM Approach Informed by Geo-Sequential Feature Engineering

Samar Saleh, Rutgers, The State University of New Jersey, Piscataway, NJ, United States, Weihong Guo

In additive manufacturing or 3D printing, precise control of thermal processes is pivotal for ensuring product quality. Overheating, a prevalent challenge, can significantly compromise the structural integrity of 3D-printed parts. This study introduces a novel predictive modeling approach for overheating in 3D printing that circumvents the direct quantification of thermal behaviors. Through advanced feature engineering, we have incorporated geo-sequential data and scanning patterns into the dataset, enriching it with structured spatial and sequential context. We propose a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model architecture that synergistically exploits spatial patterns and sequential data inherent in 3D printing processes. The CNN component is responsible for extracting spatial features, while the LSTM layers model the sequential dependencies to predict the Consolidated Overheating Ratio (COR). This model has demonstrated superior predictive performance compared to traditional methods, as evidenced by rigorous validation on a comprehensive dataset. While overheating has only been understood qualitatively in the past, the proposed approach can effectively predict overheating quantitatively without requiring explicit thermal data. This simplifies the predictive process and enables real-time corrective measures in the printing parameters. This study not only advances overheating prediction methodologies but also contributes as an expandable framework for enhancing quality control in additive manufacturing.

2 - Transfer Learning Model for Industrial Prognostics

Kyoung Min Kim, North Carolina State University, Raleigh, NC, United States

In this talk, we investigate transfer learning within the context of high-dimensional location-scale regression models. Our goal is to improve the model's accuracy on target data by using relevant information from appropriate source data. We have developed two algorithms: The first is a transfer learning algorithm for the location-scale regression model, which assumes that all sources are transferable. The second is a Transferable Source Detection algorithm that determines which sources among the available data can be transferred. Additionally, we will present experiments to demonstrate the effectiveness of these two algorithms.

3 - Heterogeneity-aware Federated Learning for Collaborative Fatigue Life Prediction in Additive Manufacturing

Anyi Li, Auburn University, Auburn, AL, United States, Jia Liu

Accurately predicting the fatigue life of additive manufactured (AM) parts from inspection and testing data is crucial for their adoption in critical applications like aerospace. However, inspection and testing specimens are too limited to build an accurate predictive model for individual manufacturers. Federated learning (FL) is a promising technique enabling privacy-preserving, collaborative learning with a network of distributed AM manufacturers. However, the data among manufacturers are often heterogenous (i.e., non-IID) due to distinct machines, process conditions, or build volume, paralyzing the FL deployment for quality control in distributed AM. To alleviate this issue, we propose a heterogeneity-aware FL model for fatigue life prediction among AM manufacturers. The proposed model mainly tackles the non-IID data problem across distributed AM manufacturers' machines by integrating data heterogeneity in the FL training model. During training, it will regulate and balance the model weights between global FL and local individual AM manufacturers' models to mitigate the non-IID constraints while maintaining their model performance for fatigue life prediction. This model can benefit manufacturers with different capabilities in accurate fatigue life prediction.

4 - A Physics-Informed SPC Approach to Machine Health Monitoring for Two-Photon Lithography

Sixian Jia, University of Michigan - Ann Arbor, Ann Arbor, MI, United States, Zhiqiao Dong, Chenhui Shao

Two-photon lithography (TPL) is an advanced additive manufacturing technique to produce three-dimensional (3D) micro- and nano-scale structures. Like any complex manufacturing processes, machine health conditions of TPL systems can significantly affect the performance and quality of the fabricated structures. As such, effective and efficient monitoring of the machine health status and timely maintenance interventions are crucial to ensure consistent and high-quality production. Presently, maintenance practice for TPL machines is heavily experience-based and not informed by the real-time machine conditions. These maintenance strategies can lead to inefficiencies and non-timely interventions, thus compromising product quality. To date, there exists very limited research on machine health monitoring for TPL. Integrating new physics informed structure dimension models with classical statistical process control (SPC) philosophy and methods, this paper develops three approaches for accurate and timely monitoring of TPL machine health conditions. A large-scale experiment consisting of six TPL process parameter combinations and six structure dimensions was repeated under two machine health conditions to provide the data for assessing the effectiveness of the proposed approaches. It is shown that all three approaches are highly effective with an accuracy of up to 100 %. Importantly, the approaches are shown to have excellent robustness and generalizability. To the best of our knowledge, this research is among the first to investigate the maintenance issue in TPL.

TC73

Regency - 704

Transforming Education Through Data Analytics

Invited Session

Data Mining

Chair: Alessandro Hill, University of Bologna, Bologna, Italy

Co-Chair: Puneet Agarwal, California Polytechnic State University, San Luis Obispo, CA, United States

1 - Course Content Mining for Academic Program Selection

Puneet Agarwal, California Polytechnic State University, San Luis Obispo, CA, United States

Prospective students have to make the challenging decision of selecting a university program that will lay the foundations of their academic and professional career. For the decision maker and support services, it is often difficult and time-consuming to match personal interests with suitable programs and their complex and vast catalogue data behind. We present the first data mining-based information system that allows students to efficiently obtain recommendations that stem from both the collective program content and their personal interests. We use machine learning to analyze all available course descriptions and extract interest topics, represented by significative keywords that characterize learning content. Underpinned by the student's individual indication of interests, a ranked shortlist of most relevant programs is computed through a statistical backtracking method. In a case study at a post-secondary institution with over 80 undergraduate study programs and more than 5,000 courses, the system is capable of providing immediate and effective decision support. A qualitative study involving 65 students demonstrates that the presented interest topics are meaningful and allow for positive effects such as serendipity, personalization and fairness. Over 98% indicated that the recommendations matched their interests and about 94% agreed to using this tool in the future.

2 - Admission Process Optimization: A Two-Stage Decision Framework

Meiling Wu, University of California, Irvine, Irvine, CA, United States

The competition to attract top candidates is intense for universities, as applicants often apply to multiple schools. While scholarship offers can serve as a powerful tool to influence applicant decisions, the utility derived from financial aid incentives is heterogeneous among candidates and constrained by budgetary limitations. Furthermore, the admission process has multiple rounds, wherein the number of offers extended in each round can impact subsequent admission decisions. To address these complexities, this paper proposes a two-stage decision framework designed to optimize admission outcomes, enabling institutions to attain targeted enrollment figures while curating a class portfolio that aligns with their desired characteristics. The first stage involves predicting the enrollment probabilities of applicants based on their individual characteristics, academic performance, extracurricular activities, demographic information, and personal statements to generate probabilistic predictions of applicant enrollment. The second stage leverages these predictions by applying optimization to maximize expected utility for the institution. By assigning weights to different attributes and preferences, the algorithm optimizes the selection process while ensuring alignment with target enrollment numbers and class composition objectives.

3 - Effectively Expanding Student's In-Class Social Networks using Analytics

Alessandro Hill, California Polytechnic State University-San Luis Obispo, San Luis Obispo, CA, United States

A strong and diverse peer network is important for students' educational, personal, and professional success. We present novel methods for augmenting students' ties during class based on social network analysis. First, we explain how optimized project group assignment can create new connections. Secondly, we show how to achieve the latter using optimized seating arrangements in arbitrary classroom layouts. We solve the underlying hard combinatorial problems, variants of the set partitioning problem and the assignment problem, to optimality as integer programs using mathematical optimization. We describe the practical implementation of our techniques and account for both student and instructor needs. In case studies including more than 300 industrial engineering students, we analyze the potential and quantify the actual impact. Finally, we report student and instructor feedback which supports the usefulness of our approaches.

4 - Gamifying Analytics Education: Using Cryptocurrencies to Enhance Learning Outcomes in Business Undergraduates

Nicolas Nunez, Centrum PUCP, Lima, Peru

Addressing low analytics skills among business undergraduates, our study used cryptocurrencies in a game-like setting to boost learning outcomes in a Data Analytics course. Our experiment involved 41 students who were each given USD 10 in cryptos. They were tasked to apply their knowledge in statistics and predictive modeling to manage their crypto assets effectively. Our aim was to prevent loss in their portfolio's value while applying the concepts. Results were promising: 85% of the students completed all their tasks, 95% reported a high level of enjoyment, and interestingly, all participants profited, coinciding with a rise in prices. Our findings show that gamified, real-world scenarios can significantly enhance educational outcomes.

TC74

Regency - 705

Equity in the Energy Transition

Flash Session

ENRE: Energy-Climate

Chair: Erin Baker, Univ of Massachusetts-Amherst, Amherst, MA, United States

1 - The cost of inclusive electrification in sub-Saharan Africa

Ayse Selin Kocaman, Bilkent University, Ankara, Turkey, Beste Akbas, Alper Yilmaz, Ozlem Karsu, Philipp Trotter, Destenie Nock

The United Nations' Sustainable Development Goal 7 aims to ensure access to affordable, reliable, sustainable, and modern energy for all. Currently, there are 770 million people without access to electricity, with 75% residing in Sub-Saharan Africa. Decision-makers in countries with incomplete electrification systems face the challenge of increasing access rates while reducing system inequality. This paper proposes a framework that integrates data-driven analysis and optimization to estimate the cost of planning with an emphasis on fairness, departing from the traditional approach of cost-minimization planning. The study achieves this by incrementally increasing electricity access and solving a cost-minimizing model with and without Gini constraints, considering both on-grid and off-grid options in the electrification process. The Gini constraint aims to mitigate disparities in electrification rates across counties within underdeveloped countries. Consequently, we identify the additional cost of incorporating regional inclusivity into electrification planning. This analysis is exemplified for select countries in Sub-Saharan Africa.

2 - Evaluating equity outcomes of low-carbon energy transitions

Ranjit Deshmukh, University of California Santa Barbara, Santa Barbara, CA, United States, Sage Kime, Measrainsey Meng, Haozhe Yang, Wen-Tien Wang

Low-carbon energy transitions, while crucial for mitigating climate change, can cause energy justice and equity impacts on historically disadvantaged communities. Evaluating equity outcomes of past and potential policies, regulations, and programs enabling a low-carbon energy transition is crucial for the transition to be just. In this talk, we will present a framework that identifies and classifies energy justice and equity metrics into dimensions of health, access and livelihood to evaluate the effects of decarbonization initiatives. We will then present results from two studies that evaluate the potential equity outcomes of future low-carbon energy transitions. The first study focuses on the health and employment impacts of the phaseout of crude oil extraction and refining in the United States and the distribution of those impacts across race, income, and historically disadvantaged status. The second study focuses on the distribution of health and employment impacts of China's low-carbon transition in the electricity sector across its provinces.

3 - Carbon Capture and Co-pollutants in a Networked Power System

Paola Furlanetto, UMass Amherst, Boston, MA, United States, Erin Baker

Carbon Capture (CC) is an interesting technology for decarbonization of many sectors, including the power system. There are, however, major uncertainties about its impacts on the emission patterns of the power grid. This happens because while CO2 emissions affect the entire planet, co-pollutants (other emissions from fossil fuel generators, like NOx and SOx) are harmful for those directly exposed. CC guarantees removal of CO2 but the data is uncertain when it comes to co-pollutants and this has particular consequences for people of color and low-income who usually live nearby the power plants.

In this work we evaluate how CC affects the stock and distribution of airborne pollutants, considering technological uncertainties and policy design. We employ a Power Flow model of a three-node, mixed-source network in which fossil fuel power plants may invest in CC via retrofit. The policies include a system-wide cap on CO2, a carbon price, and a conjuncture on US rules (45Q Tax Credit, EPA Rule 111 and Good Neighbor). We report the extent of CC adoption, the energy source mix, the airborne emissions (CO2 and co-pollutants) and the monetized damages to human health. Our work shows that co-pollutant dynamics must be considered in the first steps of CC policy design, otherwise there can be stark consequences for public health.

4 - Robust Portfolio Analysis of Electricity Investments Under Net-Zero Uncertainties

Erin Baker, Univ of Massachusetts-Amherst, Amherst, MA, United States

Policy-makers and planners are looking to make robust energy system investments under deep uncertainty and conflicting objectives. This analysis presents a robust portfolio approach to address a range of uncertainties and multiple objectives and applies it to inform U.S. electric sector decisions under deep decarbonization. Results point to many plausible pathways to reaching economy-wide net-zero emissions: six of nine portfolios considered are non-dominated. Wind and solar generation play large roles in non-dominated portfolios, while dominated portfolios tend to retain fossil fuel capacity, which can increase costs and emissions. This analysis highlights the importance of considering fuel price uncertainty and, in particular, the possibility of high natural gas prices, which can lead to high co-pollution in otherwise low-polluting portfolios. Results illustrate trade-offs between emissions and costs; as well as between CO2 and co-pollutants, which is largely due to the use of carbon removal.

5 - Climate Change Impacts on Low-Income Household Energy Costs in the US

Camilo Toruño, University of Michigan, Ann Arbor, MI, United States

Mitigating climate change impacts on low-income communities is particularly important due to historic injustices, unique vulnerabilities, and low culpability for the climate crisis [1]. This report studies climate change impacts on low-income household energy burdens in the United States. Energy burdens challenge nearly a quarter of US households, and recent scholarship shows some climate scenarios have a 10-20% increases in energy burdens in the summers of Phoenix, and New York City [2, 3, 4, 5]. This research seeks to expand the field with modeling and analysis of climate change impacts on energy burdens of low-income households across 15 U.S. cities covering all contiguous U.S. climate zones. Our results show that climate change when considered across annual seasons causes monthly energy burdens to shift from winter to summer, with an overall slight decrease in energy burdens. The energy burdens in summer increase and winter decrease under RCP 8.5 by 6.6% and 9.1% respectively, corresponding to absolute energy burdens change of about 1-1.5%pt (percentage points). For low-income households that have high energy burdens (> 10%pt), any increase can be financially unmanageable, and any decrease could provide important relief. Shifting when low-income households pay for energy can be very challenging, and with warmer summers we can expect higher demand for AC adoption, yet low-income households may struggle to afford purchasing and running air conditioning. These compounding concerns highlight the need for managing low-income household's thermal comfort, carbon emissions, as well as energy burdens.

6 - Energy and economic security analysis across multiple scales in the Grid Operations, Decarbonization, Energy and Environmental Equity Platform (GODEEEP)

Kendall Mongird, Pacific Northwest National Laboratory, College Park, MD, United States, Stephanie Waldhoff

Ambitious climate policy requires major energy sector transitions, impacting energy producers, suppliers, and consumers. Pacific Northwest National Laboratory's Grid Operations, Decarbonization, Energy and Environmental Equity Platform (GODEEEP) moves beyond *how* to achieve these major transitions to explore their impacts on energy and economic well-being at multiple scales, within a consistent modeling framework.

Understanding all of these impacts requires a framework that can simultaneously explore the effects of technological change induced by GHG mitigation policies and changing socioeconomic conditions. Many of the policy-relevant metrics cross multiple spatial, temporal, and demographic scales. The GODEEEP research has focused on modeling metrics of residential energy security by income-deciles at the state level and jobs and income impacts at the county level in Washington State.

Residential energy security results show uneven impacts across income groups and states under the net zero scenarios, relative to the reference, with the largest changes in energy burden for middle-income consumer groups. For instance, the inequality in energy burden varies both within and across states and increases in the net zero scenario, with increasing shares of income spent on energy under the Net-Zero transition scenario. However, there are also positive impacts of this transition, such as an increase in the number of power sector jobs, as states transition to clean power sources.

TC75

Regency - 706

Sustainable Transportation Operations

Invited Session

ENRE: Environment and Sustainability

Chair: Isil Koyuncu, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Strategic Deployment of Electric Vehicle Charging Infrastructure in South Korea: Location and Capacity Planning

Jin ah Yang, LHRI, Seoul, Korea, Republic of, Na Rea Cho

While electric vehicles (EVs) have garnered worldwide attention, the lack of adequate charging infrastructure remains a significant hurdle for entering the EV market. To alleviate the drivers' charge anxiety and support the South Korean government's goal of accelerating EV adoption rates, we establish the problem of determining the optimal locations as well as capacities of EV chargers installed within a planned city in South Korea. Interestingly, our study accounts for a group of government regulations governing EV charging station placement, such as the mandatory installation of at least 5% or 2% of the total number of parking spaces in new or old buildings. We identify the drivers and implications of the investment in EV charging facilities. Furthermore, we supplement our analytical results with a case study of a Korean city.

2 - Electrifying Emergency Medical Service Vehicles

Mesut Yavuz, University of Alabama, Tuscaloosa, AL, United States, Ozgur Satici, Iman Dayarian, Muting Ma, Nickolas Freeman

We address the electrification of emergency medical service (EMS) vehicles. Allowing gradual electric vehicle adoption, we (i) develop a dispatching rule to assign a vehicle from a mixed (diesel and electric) fleet of ambulances, (ii) explore the tradeoffs among cost, service level and greenhouse gas emissions, and (iii) prescribe appropriate electrification levels for EMS agencies based on their spatial and demand characteristics. With actual data spanning Alabama in a 3-year period, we show that most agencies could benefit from electrifying at least half of their fleets.

3 - A Column Generation-based Matheuristic to solve the Mixed-Fleet Green Vehicle Routing

Isil Koyuncu, The University of Texas at San Antonio, San Antonio, TX, United States, Mesut Yavuz

This study addresses a mixed-fleet green vehicle routing problem with conventional and electric vehicles. We explore the differences and similarities of route-optimization efforts and their outcomes between the two types of vehicles. Informed by the data from our exploration phase, we design three procedures to be used in a matheuristic. The developed route-based decomposition framework includes column generation and set partitioning components. We computationally experiment on testbeds from the existing literature, and thereby demonstrate that the proposed approach produces results that are comparable with the exact methods that are costly in terms of CPU time.

4 - Sustainable Urban Mobility: Promoting Mobility Equity in Micromobility Sharing Services

Wu Hao, National University of Singapore, Singapore, Singapore, Long He, Sheng Liu

This study investigates the role of subsidy and regulatory policies in promoting mobility equity within micromobility sharing services. By analyzing key performance metrics such as service availability and served demand across different urban areas, we aim to identify effective policy interventions that ensure equitable access and usage. This research offers practical insights for policymakers committed to advancing social equity and sustainability in urban mobility.

TC76

Regency - 707

Energy Systems Modeling and Optimization

Invited Session

ENRE: Other Energy

Chair: Drew Kassel, University of Texas at Austin, Austin, TX, United States

1 - Impact of experience curves on national adoption of enhanced geothermal power

Wilson Ricks, Princeton University, Princeton, NJ, United States

Enhanced geothermal power is an emerging source of clean, firm electricity generation with prospects for near-term commercialization. In this work we establish a near-term cost baseline for this technology and use an electricity system capacity expansion model with endogenous technological learning-by-doing to explore pathways to large-scale geothermal deployment in the United States over the coming decades. We find that under policies targeting a near-zero or zero-carbon electricity sector by 2050, enhanced geothermal power with an assumed 15% learning rate for subsurface development could grow to become the dominant source of clean firm power on the US grid and supply more than a quarter of total generation. With sufficient learning-by-doing through initial deployments in the resource-rich western US, enhanced geothermal could become economically competitive even in the relatively cooler eastern half of the country. Long-term deployment potential for both enhanced geothermal power and other emerging clean firm technologies is sensitive to the timing and scale of initial deployments in the 2030s, and policies that drive early deployment of these technologies can significantly reduce long-run electricity system costs.

2 - Co-optimization of hydrogen and power production to support long-term grid planning in Texas

Jerry Potts, The University of Texas at Austin, Austin, TX, United States, Sergio Castellanos

Electrolytic hydrogen has emerged as an alternative fuel source for deep decarbonization due to its lack of emissions and capacity for long-duration energy storage. The electricity requirements to satisfy hydrogen demand using electrolysis are likely to have substantial impacts on long-term grid planning. In this work, we develop an open-source capacity expansion model of Texas Interconnection - a region with strong oil and gas infrastructure as well as abundant renewable energy resources that can support large-scale hydrogen adoption - to co-optimize the regional power system and hydrogen sector. We develop a wide set of scenarios to quantify the infrastructure requirements to support hydrogen development under different decarbonization goals and policy incentives. Preliminary results of this model indicate that 44.6 GW of additional transmission capacity and 125 GW of additional renewable capacity are needed to meet the electricity requirements to produce electrolytic hydrogen in a business-as-usual scenario. These capacity requirements more than double when the model is constrained to reach net-zero carbon emissions by 2050. Additional work is underway to further characterize the investment decisions in the hydrogen supply chain in each of these scenarios and to quantify the impacts of sector coupling on the power sector. The results of this analysis elucidate the infrastructure requirements associated with industrial hydrogen demand and can guide stakeholders on how to effectively leverage energy resources and optimize investment decisions to support the development of this emerging market.

3 - PyPSA-USA: A flexible open-source energy system data model and optimization tool for the United States

Kamran Tehranchi, Stanford University, Stanford, CA, United States, Trevor Barnes, Ines Azevedo

PyPSA-USA is a flexible power system data model and optimization tool for the continental United States that can be configured to address modeling questions across resolutions. The PyPSA-USA package provides an interface to flexibly translate between nodal and zonal transmission configurations, prepare generator data, and integrate regional policy constraints with high spatial resolution and temporal scope. The package integrates with climate reanalysis datasets and geospatial terrain datasets, enabling user access to decades of renewable resource availability. We present use-cases for PyPSA-USA in transmission expansion planning and historical operational simulation for model back-casting validation.

4 - Connecting Texas To The National Grids As A Reliability Measure Against Extreme Winter Weather

Drew Kassel, The University of Texas at Austin, Austin, TX, United States, Joshua Rhodes

Grid reliability in Texas is an increasingly highlighted concern due to winter storms and heat waves that have threatened power sector infrastructure and resulted in recent blackout events and calls for demand side conservation One of the most devastating of these events was the North American winter storm, later dubbed “Winter Storm Uri” by the Weather Channel, that froze the region in February 2021. This study focuses on two potential tools which can be used to improve power grid reliability against such winter storms. These two tools are building more firm generation capacity, such as natural gas power plants, and expanding interregional transmission capacity between the Electric Reliability Council of Texas (ERCOT), the Western Electricity Coordinating Council (WECC), and the Eastern Interconnect. Our method leverages open-source modeling tools, such as PowerGenome, pyGRETA, and GenX to synthesize unique zonal grid data, construct a consolidated network of model regions, and simulate different developmental pathways of capacity expansion and operational dispatch while simultaneously optimizing cost and avoiding outage events. Our initial results indicate that it is possible to construct a reliable grid capable of weathering winter storms by building many additional power plants; however, connecting the ERCOT grid to other grids lowers both costs and emissions while providing the same level of reliability in preparation for frequent, extreme, winter storms.

5 - Flexibly-operated carbon capture power plants offer near-term mitigation opportunities with variable renewable energy

Jiacong Li, Tsinghua University, Beijing, China, People's Republic of, Tingliang Zhang, Michael Davidson, Xi Lu

Electrical power system accounts for over 40% of China's total CO_2 emissions, with 56% of electricity generated from thermal power plants in 2022. Flexibly-operated carbon capture and storage (CCS) has been suggested as an effective measure for the decarbonization of coal-fired power plants, with lower capture costs and energy penalty compared with conventional CCS. This study characterizes the interplay of flexible carbon capture power plant (CCPP) with variable renewable energy (VRE) in a large regional grid, and quantifies the system-level benefit of the synergies between flexible CCPP and VRE in environmental and economic aspects, using an improved unit commitment model. Hourly characterization suggests that flexible CCS would decrease its energy consumption in peak-demand periods, and be operated at full load in off-peak periods to take advantage of otherwise curtailed renewable electricity. A number of scenarios installing flexible CCS with different configurations are evaluated. The lowest abatement cost case indicates that installing flexible CCS with 4-h solvent storage and flue gas venting (70% average capture rate) on 10% of coal-fired plants will result in 9.81% reduction of power system total carbon emission in 2030, with a CO_2 mitigation cost of 269 CNY/t (37 USD/t). To fulfill the benefits of flexible CCPP in future power systems, both policy incentives and optimal operation are required.

TC77

Regency - 708

Optimization Model Analysis and Debugging

Invited Session

Computing Society

Chair: John Chinneck, Carleton University, Ottawa, ON, Canada

1 - Recent Advances in Debugging Tools to Diagnose Ill Conditioning in Linear Programming

Ed Klotz, Gurobi Optimization, Incline Village, NV, United States

Version 1.0 of the gurobi-modelanalyzer open source Python package became available in late 2023. The primary component of the package was a function that provided a row or column based explanation of ill conditioned basis matrices. While the package has had numerous successful use cases, potential challenges remained regarding faster computation time and smaller explanations. This talk will describe the improvements made in version 1.1 of the package that address these challenges.

2 - Ensuring Robustness in Data and Model Results

Robert Randall, Princeton Consultants, Inc., Greenville, SC, United States

Optimization models provide powerful solutions, but their effectiveness hinges on the quality of input data and the rigor of solution validation. This presentation explores methodologies and tools used by Princeton Consultants to guarantee the robustness of real-world models. We'll discuss rigorous input data validation techniques to prevent infeasibilities or model breakdowns. Additionally, we'll delve into the creation of independent solution validators that ensure model outputs consistently adhere to all specified requirements. These strategies provide confidence in model-driven decision-making.

3 - Infeasible Model Analysis In the Optverse Solver

John Chinneck, Carleton University, Ottawa, ON, Canada, Cho Ho (Peter) Lam, Zirui Zhou

Isolating an Irreducible Infeasible Subset (IIS) of constraints is the best way to analyze an infeasible optimization model. The OptVerse solver incorporates very fast algorithms for this purpose. The LP analyzer takes advantage of the presolver to isolate a small subset of constraints for conventional analysis, whether or not the presolve detects infeasibility. The MIP analyzer uses new techniques to very quickly find an IIS that includes the integrality restrictions. Experimental results are given.

4 - A New Reformulated Mixed Integer Linear Programming Model for the Unequal Areas Facilities Layout Problem

Joshua Adu Afari, Missouri University of Science & Technology, Rolla, MO, United States, Abhijit Gosavi, Robert Marley

In complex real-world layout-design problems where department sizes vary, layout optimization techniques have relied on models such as the unequal-areas facility layout (UAFL) model. This model has a mixed integer programming (MILP) nature, whose solutions with branch-and-bound/cut (B-B/C) often lead to poor values of the objective function (the material-handling cost). The literature also suggests employing an aspect-ratio criterion for this MILP, which while fitting each department into a fixed area inevitably alters the departments' dimensions, causing the dimensions to stray from their designed values. To address this challenge, this research introduces a wasted-space-penalized (WASP) model for UAFL, or WASP-UAFL model for short. Unlike the aspect-ratio or the B-B/C approach, WASP-UAFL is a reformulation of the original UAFL model that modifies the objective function and allows for predetermined and unalterable dimensions and areas for departments, while delivering low values for the material-handling cost. Our research proposes numerical solutions to the WASP-UAFL model using three meta-heuristics: backtracking adaptive search, simulated annealing, and the genetic algorithm. We will present numerical results that demonstrate how a simple reformulation of a well-known MILP leads to far superior and easily obtainable solutions in practice for layout problems of realistic dimensions.

TC78

Regency - 709

Federated Learning and Optimization: I

Invited Session

Computing Society

Chair: Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

1 - Federated Adaptive Global Pruning under Model Heterogeneity

Yijiang Li, Argonne National Laboratory, Lemont, IL, United States, Guangji Bai, Kibaek Kim

Federated Learning (FL) has gained significant interest in training AI models in a distributed computing environment benefiting from its capability to maintain the privacy of sensitive data of the participating parties. However, challenges remain in effectively handling of participating parties with heterogeneous computational power, such as edge devices. In this work, we propose a federated framework that involves an adaptive global pruning scheme to enable collaborative training of large models, such as LLMs, on parties with heterogeneous computational power.

2 - Zeroth-Order Federated Methods for Stochastic MPECs and Nondifferentiable Nonconvex Hierarchical Optimization

Yuyang Qiu, Rutgers University, New Brunswick, NJ, United States, Uday Shanbhag, Farzad Yousefian

Motivated by the emergence of federated learning (FL), we design and analyze federated methods for addressing: (i) Nondifferentiable nonconvex optimization; (ii) Bilevel optimization; (iii) Minimax problems; and (iv) Two-stage stochastic mathematical programs with equilibrium constraints (2s-SMPEC). Notably, in an implicit sense, (ii), (iii), and (iv) are instances of (i). However, these hierarchical problems are often complicated by the absence of a closed form expression for the implicit objective function. Research on these problems has been limited and afflicted by reliance on strong assumptions, including the need for differentiability of the implicit function and the absence of constraints in the lower-level problem, among others. We make the following contributions. In (i), by leveraging convolution-based smoothing and Clarke's subdifferential calculus, we devise a randomized smoothing-enabled zeroth-order FL method and derive communication and iteration complexity guarantees for computing an approximate Clarke stationary point. To contend with (ii) and (iii), we devise a unifying randomized implicit zeroth-order FL framework, equipped with explicit communication and iteration complexities. Importantly, our method utilizes delays during local steps to skip calls to the inexact lower-level FL oracle. This results in significant reduction in communication overhead. In (iv), we devise an inexact implicit variant of the method in (i). Remarkably, this method achieves a total communication complexity matching that of single-level nonsmooth nonconvex optimization in FL. We empirically validate the theoretical findings on instances of federated nonsmooth and hierarchical problems including training of ReLU neural networks, hyperparameter learning, fair classification, and Stackelberg-Nash-Cournot equilibrium seeking problem.

3 - Robust Federated Learning with Differential Privacy

Xu Chen, Columbia University, New York, NY, United States, Kibaek Kim

We study the numerical performance of robust federated learning (FL) approaches against noisy data due to differential privacy (DP). DP is a widely-used privacy-preservation technique that statistically guarantees the privacy by adding noise generated from a distribution to model/data. Employing DP in FL a trade-off between data privacy and model performance. We design and numerically experiment a robust FL model with various data set.

4 - Differentially private federated learning via inexact ADMM with multiple local updates

Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States, Minseok Ryu

Differential privacy (DP) techniques can be applied to the federated learning model to statistically guarantee data privacy against inference attacks to com- munication among the learning agents. While ensuring strong data privacy, however, the DP techniques hinder achieving a greater learning performance. In this paper we develop a DP inexact alternating direction method of multipliers algorithm with multiple local updates for federated learning, where a sequence of convex subproblems is solved with the objective perturbation by random noises generated from a Laplace distribution. We show that our algorithm provides epsilon-DP for every iteration, where epsilon is a privacy budget controlled by the user. We also present convergence analyses of the proposed algorithm. Using MNIST and FEMNIST datasets for the image classification, we demonstrate that our algorithm reduces the testing error by at most 31% compared with the existing DP algorithm, while achieving the same level of data privacy. The numerical experiment also shows that our algorithm converges faster than the existing algorithm.

Tuesday PC

Flex C

Tuesday Poster Competition

Poster Session

Poster

Chair: Masha Shunko, University of Washington, Seattle, WA, United States

Co-Chair: Sripad Devalkar, Indian School of Business, Hyderabad, India

Co-Chair: Shengfan Zhang, University of Arkansas, Fayetteville, AR, United States

1 - Multi-Objective Optimization Approach for Emergency Evacuation with Pandemic Effects

Razieh Khayamim, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Ren Moses, Eren Ozguven, Maxim Dulebenets

Different types of hazards occur quite often in different parts of the globe. These hazards may cause significant property damages, monetary losses, and human fatalities. For certain types of disasters, populations are expected to evacuate from locations that anticipate the greatest impact. Emergency evacuation is generally challenging, since many people have to evacuate a given area in a short span of time. Lack of proper planning may result in negative externalities (e.g., congestion on evacuation routes, anxiety of evacuees). Furthermore, a frequent occurrence of pandemics makes emergency evacuation planning even more challenging. Rushing to the closest emergency shelter may not be the best choice, because the closest shelters may operate at the capacity level. Overcrowded emergency shelters are expected to have a high risk of virus transmission under pandemic settings. Therefore, this study proposes a new bi-objective optimization model for emergency evacuation planning, aiming not only to minimize the total travel time of evacuees to the assigned emergency shelters but also to minimize the risk of virus transmission in the assigned emergency shelters. A custom multi-objective optimization algorithm is developed to solve the proposed bi-objective optimization model. Various case studies are conducted to demonstrate applicability of the proposed methodology for real-life emergency evacuation scenarios. The findings from this research can be used to better prepare populations for approaching natural hazards and ensure their safety throughout the evacuation process. Moreover, this project will assist with minimizing virus transmission rates in emergency shelters and alleviating the negative impacts of pandemics.

2 - Advancing Carbon Mitigation via U.S. Municipal Green Bond Issuance: Insights from Causal Machine Learning

Dan Li, university of michigan, Ann Arbor, MI, United States

The green bond market provides a vital channel for directing financial resources toward climate-related initiatives that support the Sustainable Development Goals. This study assesses the impact of U.S. municipal green bonds on carbon emissions from 2009 to 2019, considering the complexities of local socio-economic conditions. Employing causal machine learning models, we discerned the effects of increased green bond issuance volume on emission reduction on county level. Our findings reveal notable heterogeneity across regions, with the lagged impact on emission reduction most conspicuous two years post-bond issuance. Within the subgroups analyzed, those residing in counties characterized by long commute times to work appear to benefit most significantly. During the urbanization process, regions experiencing prolonged commute times as a result of city expansion may observe a notably larger impact from green bond issuances in mitigating CO2 emissions. Moreover, our findings highlight the effectiveness of certified green bonds, which undergo environmental outcome verification by assurance companies, in catalyzing efforts to reduce emissions. Last but not least, our study reveals that the introduction of new green bond issuances also tends to result in emission reduction. However, the magnitude of this effect varies significantly compared to the increase in green bond volume in counties already engaged in such issuances. Our research provides novel insights that can inform the development of interventions aimed at strategically allocating resources to regions most likely to derive benefits from targeted environmental measures and urban planning initiatives.

3 - Link Prediction Modelling in Human Trafficking Networks - A Survey

Hasini Balasuriya, University of Louisville, Louisville, KY, United States, Monica Gentili

Link prediction models are vital for analyzing complex networks by predicting missing or future connections and are particularly significant in understanding and combating illicit trafficking networks. Despite numerous surveys on general link prediction modeling, there is a lack of comprehensive studies focusing on the applicability of these models to illicit networks. This study presents the findings from an extensive literature review aimed at bridging this gap by examining the effectiveness of link prediction models in identifying missing links within illicit networks. The objective is to provide a precise, well-organized overview that enhances our understanding of these models' potential in combating illicit activities, and to present preliminary results from the comparison of existing models for link prediction on a human trafficking dataset.

4 - Optimal In-Home Healthcare Shift Scheduling

Evelyn Arrey, The Ohio State University, Columbus, OH, United States, Theodore Allen, Chen Chen

We report on a scheduling app that provides optimal shift assignments for in-home healthcare workers. The problem involves substantial complexity due to various training requirements, legal restrictions, and high variability in both demand and supply with irregular hours. Moreover, in-home healthcare services involve significant face-to-face interactions, and so patients' and employees' subjective preferences are a major factor. Most organizations deal with such complexities by determining assignments manually with schedulers; however, this requires substantial training, involves high time cost in developing schedules, and elevated costs due to typically suboptimal solutions. Our solution involves mathematical optimization, deployed via a mixed integer programming model. The model is inspired by a real-world problem of helping to schedule health aides to reduce overtime expenses and achieve equitable schedules in a non-profit organization; this involves 50 weekly shifts across 40 locations and 30 employees. Our pilot test suggests that our software can reduce much of the overtime costs present in manually configured schedules. We also discuss the scheduling app that integrates our optimization model into the existing system.

5 - Advancing Real-time Pandemic Forecasting Using Large Language Models: A COVID-19 Case Study

Hongru Du, Johns Hopkins University, Baltimore, MD, United States, Jianan Zhao, Yang Zhao, Shaochong Xu, Xihong Lin, Yiran Chen, Lauren Gardner, Hao Frank Yang

Forecasting the short-term spread of an ongoing disease outbreak is a formidable challenge due to the complexity of contributing factors, some of which can be characterized through interlinked, multi-modality variables such as epidemiological time series data, viral biology, population demographics, and the intersection of public policy and human behavior. Existing forecasting model frameworks struggle with the multifaceted nature of relevant data and robust results translation, hindering their performances and providing actionable insights for public health decision-makers. Our work introduces PandemicLLM, a novel framework with multi-modal Large Language Models (LLMs) that reformulates real-time forecasting of disease spread as a text reasoning problem, with the ability to incorporate real-time, complex, non-numerical information—such as textual policies and genomic surveillance data—previously unattainable in traditional forecasting models. Through a unique AI-human cooperative prompt design and time series representation learning, this approach encodes multi-modal data for LLMs. By redefining the forecasting process as an ordinal classification task, PandemicLLM yields more robust and trustworthy predictions, facilitating public health decision-making. The model is applied to the COVID-19 pandemic and trained to utilize textual public health policies, genomic surveillance, spatial, and epidemiological time series data, and is subsequently tested across all 50 states of the U.S. for a duration of 16 weeks. Empirically, PandemicLLM is shown to be a high-performing pandemic forecasting framework that effectively captures the impact of emerging variants and can provide timely and accurate predictions. The proposed PandemicLLM opens avenues for incorporating various pandemic-related data in heterogeneous formats and exhibits performance benefits over existing models.

6 - High-resolution Hotspots of Direct and Indirect Food Instability in SSA.

JUNREN WANG, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Megan Konar, Sitian Xiong

Consistent and stable food supplies are important for ensuring food security in Africa. Climate shocks, varying trade policies, and urbanization create pressures that make stable supplies more of a dream than a reality. In this study, we focus on the specific question, 'Where does the instability come from?' We propose a framework to quickly evaluate food reliability, taking into consideration local demand as well as both domestic and international supply. The scope of this study, analyzing domestic supply at a small administrative level and assessing international contributions on a country scale, is suitable for policy decisions. This work can aid decision-makers in making informed choices regarding strategic food reserves or food aid.

7 - Strategic Optimization of Carbon Credit Incentives: Implications for Urban Travel Patterns and Emission Reductions

Dongyang Zhen, University of Maryland, College Park, College Park, MD, United States, Qingbin Cui

The transportation sector is a major contributor to global carbon emissions, accounting for a significant proportion of the overall environment footprint. In urban areas, the challenge is particularly acute, with dense population and high traffic volumes exacerbating air quality and contributing to climate change.

To address this issue, we leverage carbon credit-based incentive system as a mean to reward reductions in carbon emissions, further influence the travel behaviors. In the system, we develop a bilevel nonlinear programing model to encourage low-emission travel behaviors through the strategic use of carbon credits. The upper level, representing the decision-making of transportation authorities, aims to minimize carbon emissions through the optimal allocation and pricing of carbon credits, along with subsidies for sustainable transportation modes. The lower level reflects the utility maximization of individual travelers, incorporating travel time, cost, carbon credits, capacity constraints, and social influences into their decision-making processes. Given the nondifferentiability of the problem, quasi-Newton methods and stochastic gradient methods are employed to find the approximate solutions, as well as providing robust convergence properties and efficiency in computation.

The model's efficacy and practical implications are tested through numerical experiments utilizing historical trip data from IncenTrip. These experiments investigate the model's application across varied scenarios, reflecting different budget limits, commuter preferences, and urban settings. The experiment results reveal an inherent dilemma between travel time and emission reductions, illustrating the complex interplay between incentivizing sustainable travel and maintaining travel efficiency.

8 - A Digital Twin Model for Sorghum Growth Simulation and Breeding Optimization

Zheng Ni, Oklahoma state university, Stillwater, OK, United States, Maria Salas Fernandez, Lizhi Wang

The increasing demand for resilient and high-yielding crops necessitates advanced breeding methods. This study presents a digital twin model to simulate phenotypic data and provide breeding recommendations based on Genotype x Environment x Management (GxExM) effects for sorghum, a versatile cereal crop with varieties such as Grain (G), Forage (F), Dual Purpose (DP), and Photoperiod Sensitive (PS). The model integrates extensive environmental data, including seven years of hourly weather records and soil conditions from three locations, alongside management practices and pedigree information from 665 fathers and 141 mothers. This enhances the simulation process by analyzing 5,200 records of plot-wise final yield, 6,219 records of individual plant height, and 580 records of dry biomass at different growth stages. Additionally, the model simulates the hourly dry weight of sorghum's leaves, stems, and roots, and predicts final yield based on management practices. Demonstrating robust predictive capabilities, the digital twin model achieves a testing Relative Root Mean Squared Error (RRMSE) of 8% to 10% across diverse conditions. Within our active breeding framework, the model improves the maximum recovery rate by approximately 9% based on simulated genotype input over just two generations, compared to traditional selection strategies. By employing this innovative approach, the research identifies high-yielding hybrids, reducing the time and resources required by conventional methods. This supports efficient hybrid selection and contributes to sustainable agricultural productivity, particularly in identifying high-performing hybrids.

9 - ERGM

Minghua Zhang, University of Wisconsin, Madison, Madison, WI, United States, Justin Boutilier

The private health sector is integral to solving the Tuberculosis epidemic in low-and-middle-income countries. Private practitioner engagement programs can improve tuberculosis outcomes but uncertain data on referral networks and limited resources for engagement require careful selection of practitioners. We propose exponential random graph models and optimization to tackle this issue.

10 - Leveraging Network Science for Optimal Online Vehicle Dispatch on Ride-hailing Platforms

Ruiting Wang, University of California, Berkeley, Berkeley, CA, United States, Jiaman Wu, Scott Moura, Marta Gonzalez

Ride-sharing reduces carbon emissions by lessening private car use. The efficiency at which these platforms reduce emissions hinges on the design of their vehicle dispatch algorithms. This study aims to enhance the request service rate between drivers and riders in ride-hailing platforms by utilizing network science metrics in the shareability network structure. In this work, we first introduce a novel network science metric known as sink proximity within the context of network flow problems, which incorporates future information on the network structures. We then show that sink proximity calculation can be approached as a problem of finding the single source longest path within a directed acyclic graph, thereby addressing the computational complexity of this measure. Last, we develop a Network Science-aided Vehicle Dispatch Algorithm (NetSciVDA) that leverages network science metrics, which effectively addresses vehicle dispatch problems encountered by ride-hailing platforms. Numerical studies demonstrate that NetSciVDA represents a significant advancement in optimizing the dispatch process through the application of network science principles.

11 - Impact of Power Outages Depends on Who Loses It: Equity-Informed Grid Resilience Planning via Stochastic Optimization

Gizem Toplu-Tutay, University of Texas-Austin, Austin, TX, United States, John Hasenbein, Erhan Kutanoglu

This research presents a two-stage stochastic optimization model for flood mitigation investments in the transmission grid, focusing on substation hardening to enhance power grid resilience against extreme weather events. It develops equity metrics based on the affected population and the duration of well-being loss to address the uneven impacts of prolonged power outages on different communities, considering social vulnerabilities. The study also incorporates a justice model inspired by the U.S. government's Justice40 Initiative.

We solve the DC optimal power flow (DC-OPF) problem for each second-stage scenario, using realistic hurricane-induced flood scenarios and a realistic Texas grid for the case study. This study illustrates the tension between minimizing power loss and improving equity among different communities. The results highlight the importance of using a composite objective to enhance the resilience of both the power grid and of disadvantaged communities through investment prioritization and electricity provision. Our research, enriched by a comprehensive sensitivity analysis, offers valuable insights for policymakers, grid operators, and utilities aiming for a more resilient and equitable power grid.

12 - Binary Gaussian Copula Synthesis: A Novel Data Augmentation Technique to Advance Machine Learning for Early Prediction of Dialysis Among CKD Patients

Hamed Khosravi, West Virginia University, Morgantown, WV, United States, Srinjoy das, Abdullah Al-Mamun, Imtiaz Ahmed

The Center for Disease Control estimates that over 37 million US adults suffer from chronic kidney disease (CKD), yet 9 out of 10 of these individuals are unaware of their condition due to the absence of symptoms in the early stages. It has a significant impact on patients' quality of life, particularly when it progresses to the need for dialysis. Early prediction of dialysis is crucial as it can significantly improve patient outcomes and assist healthcare providers in making timely and informed decisions. However, developing an effective machine learning (ML)-based Clinical Decision Support System (CDSS) for early dialysis prediction (i.e., 90 days before initiation) poses a key challenge due to the imbalanced nature of data (i.e., with far fewer cases of dialysis patients compared to non-dialysis CKD patients). This data imbalance hinders the ability of ML models to accurately predict early dialysis. To address this challenge, this study evaluates various data augmentation techniques to understand their effectiveness on real-world datasets. We propose a novel approach named Binary Gaussian Copula Synthesis (BGCS). BGCS is tailored for binary medical datasets and excels in generating synthetic minority data that mirrors the distribution of the original data. BGCS enhances early dialysis prediction by outperforming traditional methods in detecting dialysis patients. It consistently leads to superior performance across models, significantly improving over real data. For the top-performing ML model, Random Forest, BCGS achieved a 72% improvement, surpassing the state-of-the-art (SOTA) augmentation approaches.

13 - SARS-CoV-2 Infection is Associated With An Increase In New Diagnoses of Schizophrenia Spectrum and Psychotic Disorder: A Study Using the US National COVID Cohort Collaborative (N3C)

Ahmed Shoyeb Raihan, West Virginia University, Morgantown, WV, United States, Imtiaz Ahmed

Amid the ongoing global repercussions of SARS-CoV-2, it is crucial to comprehend its potential long-term psychiatric effects. Several recent studies have suggested a link between COVID-19 and subsequent mental health disorders. Our investigation joins this exploration, concentrating on Schizophrenia Spectrum and Psychotic Disorders (SSPD). Different from other studies, we took acute respiratory distress syndrome (ARDS) and COVID-19 lab-negative cohorts as control groups to accurately gauge the impact of COVID-19 on SSPD. Data from 19,344,698 patients, sourced from the N3C Data Enclave platform, were methodically filtered to create propensity matched cohorts: ARDS (n = 222,337), COVID-19 positive (n = 219,264), and COVID-19 negative (n = 213,183). We systematically analyzed the hazard rate of new-onset SSPD across three distinct time intervals: 0-21 days, 22-90 days, and beyond 90 days post-infection. COVID-19 positive patients consistently exhibited a heightened hazard ratio (HR) across all intervals [0-21 days (HR: 4.6; CI: 3.7-5.7), 22-90 days (HR: 2.9; CI: 2.3 -3.8), beyond 90 days (HR: 1.7; CI: 1.5-1.)]. These are notably higher than both ARDS and COVID-19 lab-negative patients. Validations using various tests, including the Cochran Mantel Haenszel Test, Wald Test, and Log-rank Test confirmed these associations. Intriguingly, our data indicated that younger individuals face a heightened risk of SSPD after contracting COVID-19, a trend not observed in the ARDS and COVID-19 negative groups. These results, aligned with the known neurotropism of SARS-CoV-2 and earlier studies, accentuate the need for vigilant psychiatric assessment and support in the era of Long-COVID, especially among younger populations.

14 - Exploring Nutritional Equity in Foodbank Supply Chains

Nowshin Sharmile, North Carolina A&T State University, Greensboro, NC, United States, Lauren Davis

Food insecurity, defined as insufficient access to food for a healthy and active life, affected approximately 12.8% of U.S. households in 2022. A concerted effort from both the government and non-government organizations is underway to address this challenge in the United States. This study centers on the Foodbank of Central and Eastern North Carolina (FBCENC), a nonprofit hunger relief organization pivotal in collecting and distributing food donations to local agencies serving individuals in need. However, despite the critical role of food banks, nutritional considerations are often overlooked. To address this gap, the study employs the Healthy Eating Research (HER) Nutrition Guideline, categorizing nutrition types (Red, Yellow, and Green) to assess and enhance the nutritional equity of the current distribution system. A linear programming model is proposed, and equity and effectiveness are adopted as the performance measure. The study aims to develop a model strategically reducing nutritional disparities across the network. By incorporating HER guidelines and emphasizing equity in distribution, this research contributes to the broader objective of creating a more nutritionally equitable response to food insecurity within the non-profit sector.

Tuesday Poster

Flex C

Tuesday Poster Session

Poster Session

Poster

Chair: Masha Shunko, University of Washington, Seattle, WA, United States

Co-Chair: Shengfan Zhang, University of Arkansas, Fayetteville, AR, United States

Co-Chair: Sripad Devalkar, Indian School of Business, Hyderabad, India

1 - Integration of Object Detection Machine Learning and Geospatial Information for Improved Power System Planning

Maxwell Miller, Carleton University, Ottawa, ON, Canada, Ahmed Abdulla

Decarbonization of transportation and heating requires electrification, entailing significant upgrades to electricity distribution infrastructure. It is rare for local distribution companies to make publicly available maps of their systems that enable power flow simulations at the distribution level. Here, we map local distribution networks using a combination of machine learning models and geospatial intelligence. We leverage object detection machine learning models, an area of growing interest and work. Current models claim an F1 score of 0.764; however, these models focus on mapping much larger transmission infrastructure. The challenge of mapping smaller distribution infrastructure remains unresolved. This project develops widely applicable methods to improve F1 scores when mapping local distribution networks using data-filtering (focusing on feasible pole locations), and context maps. The latter integrate real-world logic with model predictions by weighting pole probabilities based on their relative positions to landmarks like roads and houses. Using Cornwall, Ontario, Canada as a test site, mapping makes use of DRAPE remote sensing imagery segmented with ArcGIS as training data for a python-based machine learning model. The model outputs a geospatial inferred dataset of distribution poles, tuned with a human-in-the-loop process. By combining this map with energy demand simulations developed elsewhere, power system optimization at the distribution level is possible. Future research will focus on power system optimization, accounting for supply at the distribution level with increased granularity and generalizing the analytical framework for other uses.

2 - The *FastMap* Redistricting Algorithm: Promoting Fair Elections with Optimized Political Districts

Matthew Petering, District Solutions LLC, Milwaukee, WI, United States

On January 12, 2024 seven proposals for Wisconsin's legislative districts were submitted to the Wisconsin Supreme Court in the case *Clarke v. Wisconsin Elections Commission*. One was generated by a computer algorithm, and six were created by expert human mapmakers. In this presentation, we examine a major development from the case: the map proposal generated by the *FastMap* algorithm significantly outperformed the other proposals. Relevant background information, an algorithm summary, and map visualizations/comparisons are provided. The case likely marks the moment when computer algorithms surpassed humans in overall mapmaking ability. Going forward, algorithmic mapmaking will enhance American democracy by ensuring that congressional, state legislative, and local election districts better reflect constitutional requirements, the will of the voters, and other criteria specified by courts and legal teams.

3 - p-Hacking and Publication Bias in Information Systems Research

Minh Nguyen, Michigan State University, East Lansing, MI, United States

p-hacking and publication bias are significant issues within the scientific community. This study examines their prevalence in quasi-experimental causal inference methods specifically within information systems research. Analyzing nearly 3,000 tests from over 300 articles published in the top seven Information Systems journals, we find notable variation in the prevalence of p-hacking and publication bias across different methods. Evidence of p-hacking is present at the 5% and 10% significance levels in Difference-in-Differences and Instrumental Variables methods, raising concerns about the reliability of results reported using these methods. Conversely, Regression Discontinuity and Synthetic Control methods do not exhibit such evidence. Importantly, our findings indicate that the review process helps address these issues, mitigating concerns about published research.

4 - A region-dependent e-hailing service pricing strategy for rapid massive evacuation

HANG SU, Tongji University, Shanghai, China, People's Republic of, XIAONING ZHANG

Faced with growing and unpredictable passenger travel demands in urban areas, e-hailing services are crucial for regional transportation. Effective pricing and scheduling strategies are urgently needed to optimize vehicle supply and enhance evacuation efficiency during mass gathering events. Based on the Macroscopic Fundamental Diagram (MFD), this paper introduces a mathematical model to characterize the time-varying evolution of e-hailing vehicles within a two-region transport system. Considering drivers' repositioning decisions and passengers' demand dynamics, the model facilitates a region-dependent pricing strategy for rapid massive evacuation. Results from the experimental study indicate that (1) the region-dependent pricing strategy consistently leads to a shorter total evacuation time, more stable traffic conditions, and reduced passenger waiting time for dispatching; and (2) despite the higher generalized wage for drivers and increased generalized cost for passengers, the region-dependent pricing strategy effectively allocates a more balanced distribution of vacant vehicle supply under mass gathering events.

5 - Does Virtual Take Longer? An Empirical Investigation of Service Time in Telemental Health

Yupu Sun, University College London, London, United Kingdom, Ersin Korpeoglu, Lina Song

Telemedicine visits, characterized by synchronous interactions between patients and providers, have witnessed a significant surge during the COVID-19 pandemic. To better understand how telemedicine impacts the efficiency of healthcare delivery, we examines the causal relationship between telemedicine adoption and service time, using the patient-level data for psychology visits from Medicare and private insurance claims in the U.S.. We employ an instrument variable model with a control function approach by utilizing the availability of 5G broadband coverage in the area as an IV. We see that telemedicine adoption significantly decreases service time in psychiatry for private insurance patients, while it significantly increases service time for the Medicare insurance case. We further find that telemedicine is consistently correlated with shorter service times for patients under 65. We posit that our observations hinge on the interaction of two opposing drivers: limited computer literacy and increasing access to care. Evidence shows that telemedicine-only patients have more monthly visits than in-person-only patients under both insurance types. The former effect drives the result for the older population whereas the latter effect drives the result for the younger population. Our study makes important contributions by providing a insight into how telemedicine use affects service time. Practitioners should be mindful of patient age when determining whether to provide virtual or face-to-face services. In particular, transforming mental health visits to online for the elderly population may put some extra strain on healthcare operations due to increased service times and visit frequencies.

6 - A multi-method approach to identifying necessary and sufficient conditions of information asymmetry in supply chain networks

Soode Vaezinejad, University of Rhode Island, Kingston, RI, United States, Dara Schniederjans

Supply chains involve various actors interconnected by finance, materials, and information flows. Information asymmetry arises when one of these actors has more or better information than another, leading to market inefficiencies and other problems. This study conducts a systematic literature review and a Delphi study with industry experts to identify conditions that cause information asymmetry in supply chains. Then, using Necessary Condition Analysis (NCA) and fuzzy-set Qualitative Comparative Analysis (fsQCA), this study identifies necessary conditions and sufficient configurations of conditions that cause information asymmetry at the interorganizational level in supply chains. This study enhances theoretical understanding and provides practical insights.

7 - Let Me Know What Appeals to You: A Flow-Inspired Bayesian Dynamic Interest Model

Jinnan Huang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Jiapeng Liu, Xiuwu Liao, Zice Ru

We introduce a novel, interpretable Bayesian dynamic interest model inspired by flow theory to understand user behavior and content consumption interests on short video platforms (SVPs). Our approach integrates a Gaussian Mixture based Hidden Markov Model (GMM-HMM) and a Topic Model within a full Bayesian framework to jointly model user behavior patterns and content consumption interests, addressing the challenge of predicting complex user interests on SVPs. This method reveals distinct user behavior stages and successfully captures the dynamic evolution of content consumption interests across these stages. It provides platform managers with interpretable insights into user interests, helping them understand the relationships between user stickiness, active states, and content preferences. To make our approach scalable to large dataset, we design an efficient variational inference algorithm for model inference. We utilize real data from NetEase Cloud Music Cloud Village, a renowned music streaming community in China, and verify the effectiveness of the proposed model through multiple comparative experiments.

8 - Optimizing Stochastic Sustainable Supply Chain Networks for Reusable Products

Amir Hossein Sadeghi, NC State University, Raleigh, NC, United States, Rob Handfield

This study addresses the problem of supply chain network design for reusable products in a single-vendor, multi-product, multi-retailer, with a focus on economic and environmental sustainability. The objective is to efficiently manage product reuse and recovery, considering various constraints such as budget limitations, storage capacity, and the number of orders. To capture the uncertainty associated with these constraints, a stochastic formulation is adopted to provide a more realistic representation of the problem. A nonlinear mathematical model is proposed to formulate the stock management problem and it is solved using SAS 9.4. This research contributes to the field of sustainability by offering a comprehensive framework for managing stock of reusable products, considering both economic and environmental dimensions. The proposed mathematical model and metaheuristic approaches, provide practical tools for companies to optimize their stock management strategies and promote sustainable practices in product reuse and recovery.

9 - On the Robustness of Spectral Algorithms for Semirandom Stochastic Block Models

Naren Manoj, TTIC, Chicago, IL, United States, Aditya Bhaskara, Agastya Jha, Michael Kapralov, Davide Mazzali, Weronika Wrosz-Kaminska

In a graph bisection problem, we are given a graph G with two equally-sized unlabeled communities, and the goal is to recover the vertices in these communities. A popular heuristic, known as spectral clustering, is to output an estimated community assignment based on the eigenvector corresponding to the second-smallest eigenvalue of the Laplacian of G. However, spectral clustering is known to be non-robust to model misspecification. Techniques based on semidefinite programming have been shown to be more robust, but they incur significant computational overheads. In this work, we study the robustness of spectral algorithms against semirandom adversaries. Informally, a semirandom adversary is allowed to ``helpfully'' change the specification of the model in a way that is consistent with the ground-truth solution. Our semirandom adversaries in particular are allowed to add edges inside clusters or increase the probability that an edge appears inside a cluster. Semirandom adversaries are a useful tool to determine the extent to which an algorithm has overfit to statistical assumptions on the input. On the positive side, we identify a wide range of semirandom adversaries under which spectral bisection using the unnormalized Laplacian is strongly consistent, i.e., it exactly recovers the planted partitioning. On the negative side, we show that in many of these settings, normalized spectral bisection outputs a partitioning that makes a classification mistake on a constant fraction of the vertices. Finally, we demonstrate numerical experiments that complement our theoretical findings.

10 - An Extension of Harsany's Linear Tracing Procedure for Selecting a Sequential Equilibrium

Lingxuan Tang, City University of Hong Kong, Kowloon, Hong Kong, Chuangyin Dang

As a strict refinement of subgame perfect equilibrium, the concept of sequential equilibrium was established by Kreps and Wilson (1982) for extensive-form games with perfect recall. The selection of an equilibrium is indispensable for applications. It is well known that Harsanyi's linear tracing procedure plays a central role in the selection of a Nash equilibrium. This paper extends Harsanyi's linear tracing procedure to selecting a sequential equilibrium by introducing an extra variable that decreases from one to zero. However, the starting point may not be uniquely determined. To address this concern, we construct a convex-quadratic-penalty agent extensive-form game by incorporating convex-quadratic-penalty terms into payoff functions of the agents. Through this formulation, we establish the existence of a smooth path that starts from an arbitrary totally mixed behavioral strategy profile and gradually converges to a sequential equilibrium as the extra variable approaches zero. Numerical results show that the differentiable path-following method is effective and efficient.

11 - Probabilistic Pruning for Branch and Bound: An example on TSP

John Becker, USNA, Annapolis, MD, United States, Rajan Batta, Claudio Contardo

This presentation introduces the probabilistic pruning method for branch and bound for reducing the time required to accept satisfactory solutions. The proposed method is tested in instances of the traveling salesman problem (TSP) to demonstrate its effectiveness. The foundation of probabilistic pruning is sampling the distribution of solutions which can occur below a given vertex of the search tree. This distribution is used to calculate the probability the incumbent solution already satisfies the gap requirement set forth by a user. If the probability is sufficiently high, the vertex is trimmed. Three variants to the pruning method are developed and tested against each other and classic branch and bound. The variants s are tested on 20-40 city TSP instances across varying probability and desired gap thresholds. The methods are also tested with a hot and cool start on each instance, probability setting, and desired gap setting. The results indicate the desired gap is almost always achieved across all probability values, and the variants can reduce search runtimes in many settings.

12 - Driving Equity Crowdfunding Success: Insights from Startup Characteristics

Emily Mo, University of California, Berkeley, Berkeley, CA, United States, Xiahua Wei

Equity crowdfunding (ECF) has emerged as a significant fundraising avenue for startups. Understanding the factors driving ECF is crucial for startup success. This study aims to identify potential factors influencing investor decisions and provide entrepreneurs with strategies to maximize their investment potential when designing campaigns. Toward this goal, we collected and analyzed campaign-level data from a leading ECF platform in the U.S., including business background information, financials, and founder demographics. Our analysis indicates that startups with higher fundraising goals, larger amounts of cash on hand, and greater net losses, as well as those with smaller valuation caps, tend to attract more funding from investors. Interestingly, factors such as the founder's gender, race, and education, along with the startup's geographic location and technological focus, demonstrated little correlation with the amount of funds raised. These findings offer insights into designing effective ECF campaigns, suggesting that entrepreneurs may benefit from setting larger funding goals and modest valuation caps. Future research will extend this study by tracking data over time, examining various stages of the fundraising process, and cross-validating findings across different ECF platforms.

13 - Multi-Objective Optimization for Shared-Ride Automated Mobility-on-Demand Services

Cheng Zhang, Texas A&M University, College Station, TX, United States, Min-Ci Sun, Luca Quadrifoglio

This study conducts a quantitative examination of shared-ride automated mobility-on-demand services (SRAMODS), highlighting potential advantages compared to non-shared rides. With the proliferation of robotaxis in certain urban areas, strategic planning and evaluating the potential for shared rides are crucial. Unlike traditional taxis, the SRAMODS system employs a centralized controller to assign vehicles, enabling a more systematic approach to vehicle allocation.

The study introduces a multi-objective model for shared-ride robotaxi, addressing the perspectives of on-site, in-vehicle riders, and robotaxi operators by minimizing waiting time, travel time, and detour distance at each pairing epoch. The proposed model can handle both ad-hoc and pre-booked riders, dynamically matching riders within each epoch, with each robotaxi accommodating up to 4 riders simultaneously.

The model is verified using conventional taxi data from Chicago. Results indicate that the weighting of waiting time, travel duration, and detour distance significantly impacts the final solutions. Prioritizing one objective improves its performance but compromises others. Equal weighting of waiting time and travel time minimizes total trip time, while balanced weighting across all objectives serves more riders. Additionally, higher vehicle capacity enhances pairing efficiency and reduces average travel distance, especially with limited robotaxis in the city.

Rider-matching models have been analyzed within the contexts of dial-a-ride, ride-hailing, and ride-sharing services. With the trend towards robotaxis, it is critical to customize models for automated mobility-on-demand services. This study investigates the incorporation of multiple perspectives, bringing attention to the benefits for all stakeholders and facilitating a smarter approach to robotaxi services.

14 - Patch-based representation learning for time-series anomaly detection

Jinju Park, Sungkyunkwan University, Suwon, Korea, Republic of, Seokho Kang

Time-series anomaly detection involves identifying anomalous patterns within time-series data, achieved by learning solely from previously observed normal patterns. Recent research advances in this field have focused on capturing long-term dependencies using large transformer architectures. Although they have proven effective, capturing long-term dependencies may not necessarily benefit anomaly detection performance in many real-world applications where time-series data appear as repetitions of regular patterns, whereas anomaly patterns break this regularity. In this study, we propose a time-series anomaly detection method based on patch-based representation learning. We use a simple representation model that takes a patch—a short segment of the time-series—as input, instead of the entire time-series. In the training phase, the model is trained using a learning objective comprising two loss functions: contrastive loss and pretext loss. The contrastive loss aims to make representations invariant to trivial noise or shifts by encouraging slightly time-shifted patches to be close together while keeping other patch pairs far apart in the representation space. The pretext loss aims to enrich representations to reflect positional relationships among patches by predicting whether a given pair of patches is adjacent. In the inference phase, whether a query is anomalous is determined based on its distance from the nearest patches in the training data and the prediction accuracy of its relative position against the timely-adjacent patch. We demonstrate the effectiveness of the proposed method on benchmark datasets.

15 - What does it take to be 100% Renewable

Eina Ooka, The Energy Authority, Bellingham, WA, United States

As wind and solar power increasingly saturate the energy grid, integrating energy storage solutions becomes essential to continue the expansion of green energy. This study utilizes linear programming to identify the optimal combination of renewable energy sources and storage systems necessary for achieving a 100% renewable energy system, its economic burden, and the benefit of load flexibility.

Using historical wind and solar data, we capture realistic energy drought situations in the US Southwest region. We evaluate multiple layers of battery storage with different durations to determine the most effective configurations. We also find that small percentages of load shedding significantly reduce the required capacities for energy generation and storage.

The results provide strategic insights into cost-effective pathways for scaling up renewable energy deployment and guide informed decision-making in achieving a sustainable and economically viable energy future.

16 - New Perspectives in Online Contract Design

Shiliang Zuo, UIUC, Savoy, IL, United States

I study the contract design problem from an online learning perspective. I study three settings. In the first setting, I study the problem of learning the optimal learning contract when the outcome is binary. I show that the problem is in fact equivalent to dynamic pricing with unknown demand curve. In the second setting, I study the multitask principal-agent problem. I first show that under certain conditions, the optimal contract is uniform across hetergeneous agents whose productivity and cost function may differ. Further, I identity the problem of finding the optimal contract as a instrumental regression problem and give efficient algorithms for learning the optimal contract. In the third setting, I study the team production setting. I introduce the notion of induced production function, and show how the principal can find the optimal profit sharing rule via solving a family of convex programs.

17 - dynamic routing and pricing for meal delivery using a deep Q learning method

Chun-wei YANG, City university of hong kong, Hong Kong, Hong Kong

On-demand meal delivery platforms require efficient routing and pricing to maximize operational efficiency and customer satisfaction. This study presents a dynamic framework for meal delivery using deep Q-learning. Our approach adapts in real-time to changing conditions and customer demands. By modeling the problem as a Markov decision process (MDP), we use deep reinforcement learning to optimize pricing and routing decisions dynamically. Our method incentivizes customers to select diffenent meal delivery options to balance load and improving efficiency. Computational experiments show our dynamic method significantly outperforms conventional strategies, increasing both revenue and orders fulfilled.

18 - Prediction-Based Traffic Signal Control Algorithm for Mixed Autonomous Vehicle Traffic Environment

Jaewon Choi, Pohang University of Science and Technology, Pohang, Korea, Republic of, Wonjae Lee, Hyungjun Park, SunJin Han, Dong Gu Choi

A robust traffic signal control system is required to accommodate the coexistence of conventional and autonomous vehicles. To achieve such system, it is necessary to predict traffic volume by considering the characteristics of autonomous vehicles. Then optimal signal decisions can be made based on the predicted traffic volumes, thereby enhancing traffic operational efficiency. We focus on developing a method to predict the traffic inflow and determine the optimal signal cycle. We perform a simulation experiment in a single intersection environment and show effective traffic management compared to the non-predictive control approach.

19 - Optimizing Inbound Logistics in Aurubis

Sofiya Onyshkevych, Aurubis AG, Hamburg, Germany, Christos Galanopoulos, Leonardo Assis, Bianca Springub

Aurubis is a leading global supplier of non-ferrous metals and one of the largest copper recyclers worldwide. Ensuring the timely receipt and processing of raw materials and components is crucial to maintaining smooth production flows. To always have enough materials available, it is essential to ensure smooth supply chain operations, including efficient planning of the unloading of the vessels and their delivery to the destinations.

To address these challenges, we have developed a data-driven optimization tool designed to support Aurubis' SCM team in daily operational planning. Our tool employs a Mixed-Integer Linear Programming (MILP) model formulated using Pyomo, incorporating various constraints on materials, unloading destinations, and different operational rules.

The primary objective of our tool is to minimize waiting times and associated costs, thereby enhancing operational efficiency. The cost function within our model accounts for multiple types of costs and incorporates detailed operational complexities. Given the intricate nature of vessel assignment and sequencing, our goal is to provide a user-friendly tool that would be used in the company regularly and efficiently.

This tool not only simplifies the assignment and scheduling processes but also optimizes the overall supply chain performance, ensuring that materials are available when needed without incurring unnecessary costs. Our presentation will delve into the development, implementation, and practical applications of this optimization tool within Aurubis' operations.

20 - Assessing Concordance of Real-Time Self-Reported Pain Events between Patients with Advanced Cancer and their Family Caregivers Using a Remote Health Monitoring System

Mina Ostovari, University of Virginia, Charlottesville , VA, United States, Natalie Crimp , Sarah Ratcliffe , Virginia LeBaron

This study proposes an analytic approach to examine the concordance (agreement) between cancer pain reported by patients and their family caregivers using real-time data collected through the Behavioral and Environmental Sensing and Intervention for Cancer (BESI-C), a remote health monitoring system designed for managing advanced cancer pain at home, particularly in rural areas. Unlike previous cross-sectional studies focusing on either patient or caregiver perspectives, this method captures dynamic symptom fluctuations and compares both viewpoints.

Patient-caregiver dyads recruited from a palliative care clinic used the BESI-C app on their smartwatches to independently record patient pain as it occurred. Patient-reported pain events were considered the reference point and clustered together if marked within a 40-minute window. Pain events recorded by a caregiver within a patient cluster or 30 minutes before or after were considered concordant with that cluster. For each pain event/cluster, we considered three scenarios: 1) Recorded by the patient with no overlapping caregiver event, 2) Only recorded by the caregiver, and 3) Recorded by the patient and overlapped with a caregiver event. The agreement was calculated by dividing scenario three occurrences by the total number of cases across all scenarios.

The sample included 12 dyads. The highest level of agreement was 60%, with all other deployments showing less than 50% agreement.

This proposed approach facilitates research on improving cancer pain management by enabling comparisons of symptom experiences between patients and their caregivers. Future work will explore the concordance relationship with quality-of-life metrics to design tailored interventions that alleviate symptom burden.

21 - Fortifying Distribution Network Nodes Subject to Disruptions

Pelin Kesrit, Texas A&M University, College Station, TX, United States, Bahar Çavdar, Joseph Geunes

We consider a distribution network for delivering a natural resource or physical good to a set of nodes, each of which serves a set of customers, in which flow disruptions may occur at one or more nodes. Each node receives flow through a path from one or more source nodes, implying that a node experiences a disruption if a disruption occurs at one or more nodes on each path from a source node. All nodes in the network are subject to a future disturbance of an uncertain degree of severity, and we assume we can quantify the degree of severity, and that it follows some well-defined probability distribution. For each node in the network, we wish to determine a fortification level that enables the node to withstand a disturbance up to some level of severity. The cost of fortification is nondecreasing in the maximum-severity fortification level chosen, and we wish to maximize the expected number of customers who do not experience a disruption following the occurrence of a disturbance, given a limited fortification budget. We formulate this problem as a mathematical program, characterize useful properties of optimal solutions, and provide methods for determining optimal fortification levels under various assumptions on the probability distribution of the disturbance severity and the network structure.

22 - Low-carbon Optimal Dispatch of Integrated Power and Transportation System via Distributed Approach

Zepeng Li, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, China, People's Republic of, Qiuwei Wu, Litao Zheng, Xuan Zhang

The rapid proliferation of electrical vehicles (EV) contributes to carbon reduction in the transportation sector. However, it also complicates the management of power distribution system (PDS) and transportation system (TS). To better manage the two systems, a low-carbon optimal dispatch scheme for the integrated power and transportation system (IPTS) is proposed. Carbon emissions from the TS and PDS are quantified using specific models. A series of model reformulation methods are employed to transform the original non-convex model into a tractable one. Then the fast adaptive alternating direction method of multipliers (ADMM) based distributed approach is developed to solve the problem to protect the privacy of the PDS and TS. Case studies illustrate the effectiveness of the proposed scheme in achieving low-carbon optimal dispatch of IPTS and the superiority of the fast adaptive ADMM based distributed approach in computation performance.

23 - The impact of Paid Membership Programs on Customer Purchases

Yuchen Xu, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Yuan Chen

Despite the widespread belief in the importance of paid membership programs, their effectiveness and profitability are still unclear. This study utilizes data from a prominent Chinese offline retailer to examine the causal effect of such programs on customer purchases. Through a quasi-experimental design incorporating a staggered difference-in-differences approach, the research investigates the impact of paid memberships on purchase amount, frequency, and average order amount. The findings show a significant positive effect on all three purchasing metrics, indicating the effectiveness of paid membership programs in increasing retail sales and customer engagement. The analysis further segments customers into low-frequency, general, discount-sensitive, and active categories, discovering that the paid membership program is particularly effective in increasing purchases among low-frequency customers. It also highlights the role of in-store services offered through offline paid memberships in affecting customer purchases. Moreover, the study explores the proxy profitability of the program and finds that each paid member can bring a positive revenue for the retailer. The study also conducts some robustness checks including parallel trend test, falsification test and causal forest algorithm as an alternative method. The research contributes valuable insights into the strategic implementation of paid membership programs in the retail sector, emphasizing their potential to enhance retailer competitiveness and profitability.

24 - Supply Chain Coordination Using Put-Option Contracts with Cooperative Advertising: A Supplier-Led Stackelberg Game

Muhammad Mudassar SHARIF, Lingnan University, Hong Kong, Hong Kong, Mingming Leng

This article deals with the coordination issue of a two-echelon supply chain facing additive promotional efforts dependent stochastic demand in the presence of cooperative advertising. Coordination becomes more challenging when the end consumer's demand is promotional dependent. As a Stackelberg follower, the retailer determines the order quantity, put-option quantity, and promotional efforts before the selling season starts. First, we proposed a put-option contract to coordinate the promotional efforts and ordering decisions of the retailer. However, we found that the put-option contract alone fails to coordinate the ordering and efforts decisions of the retailer. Instead, a put-option contract in the presence of cooperative advertising, with two parameters - option price and participation rate, can achieve supply chain coordination.

We also discussed the impact of the proposed contract on the optimal ordering and promotional effort decisions of the retailer and obtained the optimal option price and participation rate of the supplier in decentralized settings. By comparing the proposed coordinated contract with the decentralized settings and a wholesale price contract as a benchmark, we showed that the given contract improves the profits of the retailer, supplier, and leads to win-win outcomes for both the retailer and supplier.

Finally, we provided sensitivity analysis to discuss the impact of non-contract parameters and the randomness of demand on the optimal decisions, profits of the chain members, and the efficiency of the coordination and drawn some important managerial insights.

Keywords: Supply chain coordination; cooperative advertising; put-option contract; additive stochastic demand; promotional efforts; participation rate; Stackelberg game

25 - Evaluating the Potential of Generative AI for Data Storytelling: Insights from a Survey Study

Vinayaka Gude, Elon University, Elon, NC, United States, Andini Safitri, Yazan Kakish

Data storytelling synthesizes narrative techniques and data visualization to clarify complex information for diverse audiences. With the advent of Generative AI, this practice has seen substantial innovation, allowing for the automation of storytelling elements and enhancing data-driven decision-making. Today, the integration of these technologies spans text and visual content generation; however, challenges remain in accurately replicating intricate details and mitigating biases. The current study explores the strategic use of Generative AI tools in data storytelling, focusing on the optimal timing and approaches for their employment and evaluating the user's perception of the overall applicability of LLM models for the storytelling process. Our research questions aimed to understand how prior experience with data visualization and Generative AI tools influences attitudes towards their adoption and how hands-on experiences alter perceptions of their utility. Our findings suggest that individuals with prior experience using Generative AI will likely continue utilizing these tools for data storytelling, especially since they are helpful in the initial, more ambiguous stages of story development. We observed a significant shift in user likelihood to use these tools from before to after hands-on experience, highlighting the crucial role of proper education and guidance in maximizing their effectiveness. This research contributes to the ongoing dialogue on integrating advanced AI tools into additional data storytelling practices, aiming for a future where these technologies support storytelling's technical and creative aspects.

26 - Driving Sustainability - Route Optimization for Campus Composting Operations at Texas State University

Clara Novoa, Texas State University, San Marcos, TX, United States, Yeasir Iqbal, Luis Pardo, Rakesh Chowdhury

Bobcat Blend, a pioneering waste management program at Texas State University initiated in 2009. The program has expanded to manage various organic wastes, producing high-quality compost. Despite its success, Bobcat Blend faces challenges in optimizing its waste collection and transportation operations due to limited resources and a single vehicle capacity. This study develop an optimized routing strategy to enhance operational efficiency and sustainability by satisfying pickup and delivery simultaneously. The model minimizes transportation costs and maximizes efficiency by considering customers geographical distribution, stochastic demand, and vehicle capacity constraints. Computational results generated through the node-based and arc-based formulation reflect the efficiency in route design and adaptability to different demand scenarios. The optimal routing strategy enhances the program's capacity to divert organic waste from landfills, reduces environmental impact and operational costs, supports Texas State University's sustainability goals and offers a scalable model for other institutions. This study also underscores the potential of optimization models in advancing sustainable waste management practices, ensuring timely waste collection.

27 - Evaluating the Effects of Introducing Flexible Meeting Points to Ridesharing System in Large Urban Area

Cheng Zhang, Texas A&M University, College Station, TX, United States, Luca Quadrifoglio, Min-Ci Sun

Ridesharing has emerged as an effective strategy to meet the increased travel demand caused by the rapid growth of populations, and the trend towards urbanization in large metropolitan areas. However, traditional ridesharing system may increase traffic congestion, gas emissions, and passengers' discomfort due to detours for pickups and drop-offs.

To mitigate the adverse effects of travel detour, we advocate for a novel integration of flexible meeting points within the ridesharing system. This approach provides direct travel route from a pickup meeting point to drop-off locations, eliminating unnecessary travel distance. Meeting points can be located at the participants' origin and destination or within an acceptable distance from them.

Our study established flexible meeting points based on the spatial and temporal information of riders and drivers. We adopt a mixed integer linear programming formulation and simultaneously consider multiple objectives, including reducing the total mileage of travel detour, potential waiting time, and walking time between origin/destination and meeting points, while maintaining acceptable level of comfort for passengers and drivers. Simulation experiments utilize large-scale travel demand dataset from peak hours in urban districts of Chicago to validate our methodology. The experiment results show that the efficiency of transportation system is improved by introducing flexible meeting points to ridesharing system, compared to traditional taxi services and door-to-door shared ride services. This research intends to offer a strategy and detailed information for planners and policy makers to intervene by incentivizing a conversion, even if partial, of improving the system-wide ridesharing services.

28 - Multi-level Traffic-Responsive Tilt Camera Surveillance through Predictive Correlated Online Learning

Zilin Bian, New York University, Brooklyn, NY, United States

In urban traffic management, the primary challenge of dynamically and efficiently monitoring traffic conditions is compounded by the insufficient utilization of thousands of surveillance cameras along the intelligent transportation system. This paper introduces the multi-level Traffic-responsive Tilt Camera surveillance system (TTC-X), a novel framework designed for dynamic and efficient monitoring and management of traffic in urban networks. By leveraging widely deployed pan-tilt-cameras (PTCs), TTC-X overcomes the limitations of a fixed field of view in traditional surveillance systems by providing mobilized and 360-degree coverage. The innovation of TTC-X lies in the integration of advanced machine learning modules, including a detector-predictor-controller structure, with a novel Predictive Correlated Online Learning (PiCOL) methodology and the Spatial-Temporal Graph Predictor (STGP) for real-time traffic estimation and PTC control. The TTC-X is tested and evaluated under three experimental scenarios (e.g., maximum traffic flow capture, dynamic route planning, traffic state estimation) based on a simulation environment calibrated using real-world traffic data in Brooklyn, New York. The experimental results showed that TTC-X captured over 60\% total number of vehicles at the network level, dynamically adjusted its route recommendation in reaction to unexpected full-lane closure events, and reconstructed link-level traffic states with best MAE less than 1.25 vehicle/hour. Demonstrating scalability, cost-efficiency, and adaptability, TTC-X emerges as a powerful solution for urban traffic management in both cyber-physical and real-world environments.

29 - A Machine Learning-Based Method for Reliability Improvement of All-Terminal Networks

Seyyed Farid Hashemian, University of Arkansas, Fayetteville, AR, United States, Jose Azucena, Haitao Liao, Ed Pohl

Since the effort of calculating the exact all-terminal reliability grows exponentially, approximation methods are often applied. The performance of such data-driven models is highly dependent on the approximation model. In this work, we consider using a GNN model and compare it with previouse approximation models. Moreover, we implement a DRL algorithm to solve the network reliability improvement optimization problem which leverage the all-terminal network reliability estimation.

30 - Dynamic penalty-based dispatching for optimal EMS response times: A heuristic approach

Sevin Mohammadi, Columbia University, New York, NY, United States, Andrew Smyth

The dispatching process is inherently sequential and dynamic, with each decision impacting future resource availability. Simply sending the closest idle unit without considering the supply and demand in the unit's catchment area may not be optimal. This study introduces a dynamic penalty metric assigned to each idle ambulance, penalizing dispatches from regions with high demand and lower coverage. Dispatch decisions balance travel time and penalties for low-priority calls. We evaluated this approach through discrete event simulation in two experiments of varying sizes. First, a reinforcement learning agent trained in a small-scale experiment revealed that a greedy policy is myopic. Subsequent experiments showed that the heuristic policy, which balances penalties and travel time, outperformed the greedy policy. We further tested the heuristic policy on EMS responses in Manhattan, New York. The results, comparing empirical cumulative distribution functions (ECDF) of episode means for response times as well as the episode fractions of response times exceeding a certain time, indicated that the heuristic method provided more reliable and faster responses for high-priority calls. Specifically, 90\% of episodes under the heuristic policy had mean response times below 6 minutes, compared to 75\% under the greedy policy. Overall, the heuristic method demonstrated superior performance in minimizing response times and variability for high-priority emergencies.

31 - Online Optimization with Rational Agents

Owen Shen, MIT Operation Research Center, Cambridge , MA, United States, Patrick Jaillet

We analyze online optimization in dynamic environments involving rational human participants, focusing on online linear programming (OLP). We explore scenarios where buyers have access to either the complete or partial transaction history and model their optimal decisions. In the first scenario, buyers face a classification problem; we discover that the buyers achieve no asymptotic information asymmetry. In the second scenario, buyers use active learning to infer hidden information through testing orders; we analyze buyers' estimation accuracy and efficiency in different market structures. By considering the buyers' information asymmetry and estimation errors, we reveal the conditions by which the market price and direction are determined. By modeling the complexities introduced by active and rational participants, we offer a new perspective on online optimization and extend traditional OLP.

32 - Inpatient Management with Proximal Policy Optimization

Qihao Wu, The University of Hong Kong, Hong Kong, Hong Kong, Yong-Hong Kuo

Efficient inpatient management is crucial for providing good quality services to patients and enhancing patient satisfaction. The main challenge of inpatient management is the imbalance between patient demands and healthcare resources. Our research aims to address such a challenge with a Proximal Policy Optimization (PPO) framework to optimize inpatient flows. Specifically, we aim to determine, when a primary inpatient ward is full, whether the inpatient unit should hold the inpatient admission until the primary inpatient ward is available or can admit this patient to a secondary inpatient ward (which is not preferred). This dilemma lies between the costs of waiting and ward mismatches. Our PPO model dynamically makes inpatient admission decisions based on the state information including patients' profiles, ward capacities, and bed availabilities. Unlike static methods, our PPO approach adapts to continuously changing conditions of the environment to improve overflow decisions. We demonstrate the efficacy of our method by simulation experiments, where the simulation model is calibrated with real-world data. The experimental results indicate that our PPO framework can identify the “just right” optimal solution by reducing waiting and overflow costs, as compared with the baseline method and a pooling policy. Our study also discusses the practical considerations for implementing PPO in real-world large-scale applications integrating electronic health records.

33 - Novel optimization algorithm for vehicle routing

Digvijay Redekar, Arizona State University, Tempe, AZ, United States, Pitu Mirchandani

The vehicle routing problem in logistics operations involves finding continuous optimal routes in a network to visit set of destinations. These routes start from the source node such as store or delivery station, visit all destinations and return to the same source at the end of the route. The primary objective of these routes is to visit destinations while minimizing travel cost, which can include expenses for gas, electricity, associate salaries, and other resources. Our research aims to develop a novel efficient and faster algorithm to find an optimal route while minimizing travel costs. The traveling salesman problem (TSP) involves finding a route in a given network such that all nodes are visited exactly once. In complex logistics networks, finding such TSP optimal route may not be feasible. Some intermediate nodes may need to be visited more than once to complete the vehicle route. Such networks are not solvable using traditional TSP algorithms. We refer to these problems as minimizing revisits TSP (MRTSP). This research proposes a novel algorithm for complex TSP and MRTSP networks to find the optimal vehicle route that visits all destinations while minimizing travel costs. Application and performance of the proposed algorithm are analyzed with different network sizes to test optimality and efficiency. Proposed novel algorithm provides similar optimal solutions with better computational time compared to traditional TSP algorithms.

34 - Variance-reduction for Variational Inequality Problems with Bregman Distance Function

Zeinab Alizadeh, University of Arizona, Tucson, AZ, United States

We address variational inequalities with a finite-sum structure, proposing a novel single-loop stochastic variance-reduced algorithm using the Bregman distance function. Our method guarantees optimal convergence for monotone settings and improves existing results for non-monotone problems with weak Minty solutions. Numerical experiments demonstrate superior performance against state-of-the-art methods

35 - Bayesian Feasibility Determination with Multiple Constraints

Tingnan Gong, Georgia Institute of Technology, Atlanta, GA, United States, Liu Di, Junying He, Seong-Hee Kim, Yao Xie

We aim to efficiently determine the feasible region with respect to constraints on a group of unknown black-box functions that map discrete alternatives to real numbers with costly evaluations. Unlike many binary classification methods that primarily focus on developing a classifier based on a fixed size of training data, we determine which alternative to sample next to achieve high accuracy on feasibility determination with a minimal number of samples. To accomplish this, we use Gaussian processes as surrogate models and introduce a novel value-of-information acquisition function for adaptive sampling under multiple constraints. We thoroughly analyze the convergence of our proposed scheme and demonstrate its effectiveness through numerical experiments.

36 - A Hybrid Crop Modeling Approach for Predicting Biomass Production of Sorghum

Yanbin Chang, Oklahoma State University, Stillwater, OK, United States, Maria Salas-Fernandez, Lizhi Wang

Accurate predictions of crop yield are important for optimal resource management, especially in areas with variable climate and resource availability. In previous studies, researchers have explored two major modeling approaches, processed-based and data-driven methods. This paper proposes a new hybrid crop model for simulating the biomass sorghum growth process and predict the final yield. The model uses a detailed descriptive sorghum growth model to represent how phenotypes are determined by genotype, environment, management, and their interactions, while employing data-driven techniques to calibrate genotypic parameters from experimental data. Results illustrate that the hybrid crop model can achieve an accurate biomass prediction with around 7% relative root-mean-square-error with both training and test data sets. This approach has the potential to enhance the accuracy and applicability of biomass sorghum growth and yield prediction models.

37 - Systematic Review on Patient Personas

Jhanille Hurde, Binghamton University, Binghamton, NY, United States

Background

The development of patient personas is crucial for enhancing healthcare services by providing detailed and accurate representations of patient groups. This systematic review aims to analyze the current methodologies and applications of patient personas in healthcare, identify gaps in existing approaches, and explore the potential integration of advanced technologies such as machine learning (ML) and artificial intelligence (AI) to improve persona creation.

Objectives

1. To analyze previous patient persona creation methodologies.
2. To identify gaps and limitations in existing patient persona creation methodologies.

Method

A comprehensive search was conducted in electronic databases including PubMed, Web of Science, and Science Direct, resulting in 1,893 articles. After removing duplicates and applying inclusion and exclusion criteria, 24 articles were selected for review. The inclusion criteria focused on peer-reviewed articles published from 2013 onwards, centered on patient or user persona creation within healthcare.

Results

The use of qualitative data, though valuable, introduces biases that can affect persona accuracy. Despite the limitations in using personas, patient personas have been effectively utilized to meet patient needs and inform healthcare design decisions. However, gaps remain in the standardization and long-term implementation of personas in clinical settings, as well as in the performance metrics used to evaluate their effectiveness.

Conclusion

This review underscores the importance of patient personas in enhancing healthcare delivery but also identifies significant gaps and limitations in current methodologies. To address these issues, future research should focus on integrating AI and ML techniques into the persona creation process.

38 - Developing All-Hazard Risk Analysis (AHRA) to incorporate resilience into installation planning

Mary Mitchell, US Army Information Technology Laboratory, Vicksburg, MS, United States, Susan Wolters, Timothy Perkins, Shelia Barnett, Matt Swanson, John Richards

During the Business Process Reengineering (BPR) for the Virtual Toolbox for Installation Mission Effectiveness (VTIME), the All-Hazards Risk Analysis (AHRA) was developed to incorporate climate resilience and risk into installation planning products. The AHRA will replace the Installation Climate Resilience Plan (ICRP) as an analytical process integrated into all installation plans, instead of just being a static plan.

39 - Using Multi-Objective Genetic Algorithm for Food Delivery Service Path Optimization Considering Safety and Mileage

Pei-Fen Kuo, National Cheng Kung University, Tainan, Taiwan, Yu-Jung Lin, Chung-Wei Shen

Recently, online food delivery services have surged in popularity due to their convenience, yet this growth has raised significant safety concerns. Delivery drivers often lack professional driving training, and the pressure for swift deliveries encourages aggressive driving behaviors, heightening crash risks. Current navigation apps prioritize shortest or fastest routes for efficiency, neglecting safety considerations. Although some academic studies have developed routes focusing on safety or minimizing air pollution, these routes often sacrifice efficiency.

This study aims to develop a method for constructing safer delivery routes within acceptable travel distances. Utilizing real online food delivery data from 486 restaurants in Daan District, Taipei City, we created artificial Origin-Destination (O-D) demands and initially computed the shortest route for each O-D pair. Each route's crash risk was assessed using real crash data, and then Yen's algorithm was employed to generate alternative routes by replacing the most dangerous road segments with safer alternatives nearby. This iterative process formed the initial population for NSGA II, a well-known multi-objective genetic algorithm.

Simulation results indicate that compared to the shortest routes, the traditionally safest routes reduced average risk by 69% but increased average mileage by 145%. Conversely, our proposed routes reduced average risk by 59% with a more modest increase in mileage by only

47%. This study proposes a method for designing safer food delivery routes that delivery platforms could implement to mitigate accidents among their workers.

40 - A Multi-layer Network Analysis Approach to Weaponized Disinformation Diffusion and its Impact on Infrastructure Systems

Alice Nanyanzi, University of Oklahoma, Oklahoma, OK, United States, Kash Barker, Sridhar Radhakrishnan

This study delves into the dynamics of multi-layer networks representing an information layer and an underlying physical infrastructure layer. We focus on processes of diffusion (i.e., the spread of information or influence) in the information layer, and we focus on the performance of the flows in the physical infrastructure layer. We investigate how diffusion within one layer impacts performance in other interconnected layers under diverse conditions. Notably, we explore the effects of increasing the number of inter-layer edges, variations in diffusion parameters, and higher network density on physical infrastructure behavior. We integrate diffusion and network optimization models to study different characteristics, including potential points of failure and importance measures for critical components in both layers, and we develop strategies to improve robustness across both layers.

41 - Semiconductor Manufacturing Process Visualization: Focusing on Resource Groups with Significant Effect on Yield

Gyeunggeun Doh, Department of Industrial Management Engineering, Pohang University of Science and Technology(POSTECH), Nam-gu, Pohang 37673, South Korea, Korea, Republic of, Minseok Song, Deoksang Lee, Jeongwoo Seo, Changho Shin

In multistage semiconductor manufacturing process, the combinations of resources across various steps significantly affect yield. We propose a two step method for process visualization, focusing on significant resource groups. Association rule analysis is applied to discover significant resource groups. Then, resource-oriented transition system is utilized to visualize the overall process, highlighted with significant resource groups, thus offering insights to yield optimization.

42 - Maxflow-Based Integrated Network Design and Scheduling (INDS) Problems with Degraded Capacities

Xiaowei Guo, Clemson University, Clemson, SC, United States, Thomas Sharkey

We consider the problem of upgrading arc capacities in a network in order to maximize the total flow in the network over a time horizon. We are interested in determining both which arcs to upgrade (network design decisions) and when to upgrade them (scheduling decisions) while also considering the fact that the capacities of arcs undergoing their upgrading process will be degraded. We introduce a novel heuristic method based on properties of the maximum flow to solve the problem on realistic networks and compare its performance to a mixed-integer programming model for the problem. The heuristic method is tested on synthetic Manhattan Grid networks and real-world networks from Eastern Massachusetts, Anaheim, and Chicago.

43 - Integrating Fragility in Joint Power and Gas Transmission Models

Eric Taylor, Pennsylvania State University, University Park, PA, United States, Seth Blumsack

The resilience of coupled natural gas and power transmission systems during severe weather conditions is vital for the functioning of today's society. This study incorporates fragility curves into a two-stage stochastic expansion planning model to assess and enhance the resilience of the coupled systems when the location and severity of extreme weather events are both uncertain. The approach builds on existing research by introducing uncertainty in component failures through the use of a fragility curve method. Fragility curves, demonstrating the relationship between different wind intensities and likelihood of component failure, simulate how the system responds to geographically correlated failures. The model was integrated on a small public-domain gas-grid test system. Unsurprisingly, the use of fragility curves in these types of planning models leads to fewer network expansions. We find that even in the worst-case contingencies the chosen network expansions do not substantially reduce unserved energy. Our work suggests a need for improved planning frameworks that can handle multiple sources of uncertainty.

44 - BCFR: Incorporating "Resilience" into "Better, Cheaper, Faster" for Supply Chain Planning

Shuting Peng, Boston University, Boston, MA, United States, Sung Bo Ma, Po-Wei Chang

The increasing frequency and severity of global supply chain disruptions have motivated the development of a decision-making framework for procurement managers aiming to incorporate resilience into supplier evaluations. The methodology combines expected value analysis, indifference probabilities, and pairwise comparisons to rank suppliers based on four key dimensions: quality (better), cost (cheaper), delivery (faster), and disruption risk (resilience). Its primary contribution lies in quantifying each dimension on the same economic scale, considering both direct impacts as well as upstream and downstream ripple effects. A user interface was developed with the following inputs for each supplier: procurement cost, non-conformance rate, late delivery prediction, and disruption probability over the planning period. The framework consists of four analysis stages; each stage uses color-coded matrices to display uncertainty and binary searches to determine indifference probabilities. First, each pair of suppliers is evaluated to determine the difference in their quality-related costs, which is added to the difference in their procurement costs. Second, each pair of suppliers is evaluated to determine the difference in their delivery-related costs, which is added to the previous results. Third, each pair of suppliers is evaluated to determine the difference in their disruption-related costs, which is added to the previous results. Fourth, the Analytic Hierarchy Process (AHP) converts all of the pairwise comparisons into a ranking of suppliers. The poster provides an example of how the best supplier is chosen among a group of globally dispersed suppliers.

45 - Topology Preserving Fitting of Trimmed NURBS CAD Model to Deformed Solids

Lijie Liu, Iowa State University, Ames, IA, United States, Adarsh Krishnamurthy, Stephen Holland, Qing Li

The shape details of manufactured components are critical for next-generation cyber manufacturing systems. However, the design intent cannot generally be linked to the as-manufactured part because of geometry mismatch. The overarching goal of this research is to establish a systematic mapping framework between as-designed models and as-built geometries (which usually deviate from the design) for applications in next-generation manufacturing systems. Our method defines mappings that make it possible to leverage the design information in the context of the as-built part, which is otherwise impossible with simple reverse engineering. Our proposed work will facilitate data mapping from different resources and modalities (e.g., dimensional metrology and nondestructive evaluation methods such as computed tomography and thermography) in manufacturing and across the product lifecycle. Hence our work will enable Manufacturing 4.0, especially digital twin systems that rely on information exchange between digital models and physical parts.

46 - Optimal Valuation of Electricity and Carbon Storage Facility for Sustainability

K Jo Min, Iowa State University, Ames, IA, United States, Samantha Bradley, Eli Sallis, Nazia Nur, Mohammad Sadat, John Jackman

Towards a sustainable world, the storage of electric power as well as of carbon has recently been playing an increasingly critical role in the areas of energy and environment. In this poster presentation, we focus on the financial viability of such facilities when the electric power and carbon prices fluctuate with significant volatility. In order to understand the financial viability, we first gather relevant price details of the electric power and carbon as well as of the storage facilities. Based on such relevant information, we next show that the fluctuations of the prices closely follow a Geometric Brownian motion process, and then explain how a stochastic optimal control problem can be formulated given such a process. The resulting second order stochastic differential equation is solved for an analytical solution, leading to a threshold price that determines the financial viability of such facility under the current price information. We then apply our model and solution to a couple of empirical cases of electric power and carbon storage facilities. The preliminary results show that the electric power storage facility seems financially justified while the carbon storage facility still needs additional public support before such facility is financially viable.

47 - Salary Insights for Data-Driven Roles

Xiahua Wei, University of Washington, Bothell School of Business, Bothell, WA, United States, Emily Nguyen

The period from 2020 to 2024 marked a critical transition in the global economy due to the COVID-19 pandemic. This era redefined corporate hiring practices, notably through the widespread adoption of remote work. Following a significant job loss in 2020, the technology sector experienced a hiring surge during 2021-2022, only to face mass layoffs and a shift towards a return-to-office policy by 2023. This led to the emergence of a hybrid work model in the post-pandemic landscape. Currently, the job market remains challenging for many job seekers.

This backdrop raises critical questions about the impact of flexible work models on employee compensation and how this impact may vary across different job characteristics. Such inquiries are particularly pertinent in data-intensive industries, which are a significant part of the economy.

This study presents an analysis of salary data for data-centric roles from 2020 to 2024, including 15,614 data points across various metrics such as job title, remote work ratio, experience level, employment type, and company size. The findings indicate that experience level and job title significantly influence salary outcomes. On average, hybrid roles offer lower salaries compared to fully remote or on-site positions. Salary offerings from large and medium-sized companies are comparable, whereas small companies tend to offer the lowest average salaries.Despite the volatile job market and numerous layoffs, the demand for data-related positions continues to rise annually, with salaries increasing correspondingly. The market is also witnessing the introduction of new job titles, particularly those specializing in artificial intelligence.

48 - A Robust Optimization Model of Using Natural Gas for Electricity Generation in Saudi Arabia

Fahad Almutairi, The george washington university, Washington DC, DC, United States

This paper reviews the existing model of replacing crude oil with natural gas for power plants in Saudi Arabia. To reduce GHG emissions, SA announced the target of the energy sources by 2030. The electricity will be generated by natural gas and renewable energy. Currently, most of the turbines rely on crude oil. In contrast, the global model for the countries that export natural gas to the markets has another approach. They supply the gas capacity for local demand and global markets parallelly. SA has not exported gas globally yet until they cover the local demand and the electricity transformation. In this work, A robust optimization model for electricity generation in SA will be conducted. The suggested model will provide the optimal time frame of replacement crude turbines with gas. Achieving CO2 reduction requires a trade off with the cost and this is the objective of the model. This paper will contribute to the literature by providing an optimal model that aims to achieve a GHG emissions reduction target with the lowest cost in a specific time interval.

49 - Closing the Gaps: Optimality of Sample Average Approximation for Data-Driven Newsvendor Problems

Shilin Yuan, Tsinghua University, Beijing, China, People's Republic of, Jiameng Lyu, Bingkun Zhou, Yuan Zhou

We study the regret performance of Sample Average Approximation (SAA) for data-driven newsvendor problems with general convex inventory costs.
In literature, the optimality of SAA has not been fully established under both α-global strong convexity and (α,β)-local strong convexity (α-strongly convex within the β-neighborhood of the optimal quantity) conditions.
This paper closes the regret gaps for both conditions by proposing several new analysis techniques.
Under the (α,β)-local strong convexity condition, we prove the optimal regret bound of $\Theta(\log T/\alpha + 1/ (\alpha\beta))$ for SAA. This upper bound result demonstrates that the regret performance of SAA is only influenced by α and not by β in the long run. This insight enhances our understanding of how local properties affect the long-term regret performance of decision-making strategies.
Under the α-global strong convexity condition, we demonstrate that the worst-case regret of any data-driven method is lower bounded by $\Omega(\log T/\alpha)$, which is the first lower bound result that matches the existing upper bound with respect to both parameter α and time horizon T. This lower bound result advances the theoretical understanding of the performance limit of data-driven methods in newsvendor problems.
Moreover, our new analysis techniques for upper bound and lower bound might be of independent interest for broader data-driven problems.

50 - Analyzing and Optimizing the Delay Resilience of HSR Network under Unforeseen Events

JIN LUO, Tongji University, shanghai, China, People's Republic of, Yuling Ye, Ziyue Zhu, Yihan Tian, Wentao Zhou, Mingchu Han

General unforeseen events impair the capacity of networks temporarily, often necessitating measures such as speed limitation or service outage to overcome these events, consequently resulting in delays of trains. Enhancing the capability of high-speed railway (HSR) network to recover from delays caused by general unforeseen events is one of the focal points in transportation organization. This article aims at addressing delays of trains that resulting from general unforeseen events in HSR network. Based on analyzing the concept of resilience, the connotation of delay resilience in HSR networks is proposed. Evaluation indicators that assess the capability of HSR network to recover from delays caused by general unforeseen events are established from the perspectives of both degree and speed of recovery. Furthermore, an optimization framework for delay resilience in HSR network through operational adjustments is explored. Firstly, a multi-objective mixed integer programming (MIP) model is developed to optimize delay resilience on the fault section. Secondly, considering delays of trans-regional trains will affect the operation of adjacent sections, an optimization model of delay resilience in adjacent sections is constructed. The optimization framework is applied in the case of a local HSR network in East China. Finally, the spatial and temporal propagation characteristics of delays in HSR network under general unforeseen events are discovered, which provides insights for future operational adjustments.

51 - Master Production Scheduling Problem with Multiple Forecast Information

pouria arsalani, University of Alabama, Tuscaloosa, AL, United States, Iman Dayarian, Youngsoo Kim

Master Production Scheduling plays a vital role in manufacturing planning and control, orchestrating production activities to meet customer demand. The solution quality of this problem can significantly influence the total cost, schedule instability, and service level of a production inventory system. In this paper, we examine MPS within a rolling-horizon framework, where forecast demand data is updated at each iteration, leading to non-stationary demand patterns. Unlike previous works in this field, which consider a single estimation point for demand, this paper takes into account several forecast estimations made at different times before the actual realization, alongside various forecasting methods. The problem is modeled using stochastic dynamic programming approaches and is solved as a scenario-based stochastic model to minimize inventory and production costs. Furthermore, we evaluate and compare various forecasting methods used by our industry partner to forecast demand, incorporating different sources of information. This analysis will provide insights into how to effectively use different forecasting methods to leverage their information.

52 - A Global Supply Chain Network Equilibrium Model Considering Entrepot Trade

Rumei JI, Tongji University & The Hong Kong Polytechnic University, Shanghai, China, People's Republic of, Xiaoning Zhang, Min Xu

The global supply chain plays a vital role in worldwide economic development, which urges careful attention to the impact of international trade friction on this interconnected system. Moreover, entrepôt trade is a mitigation strategy in response to such trade tensions. However, few papers have attempted to incorporate entrepôt trade into the analysis of how international trade friction affects the global supply chain. To bridge the research gap, this paper proposes a global supply chain network game model considering entrepôt trade to investigate the influence of international trade friction on global supply chains and the role of entrepôt trade in it. The governing equilibrium conditions of the model are formulated as a variational inequality problem. Additionally, this research performs numerical experiments with a global supply chain network of one final product requiring two components to illustrate how international trade friction impacts the global supply chain and how entrepôt trade mitigates the effect. The results show that international trade friction would lead to a decline in the direct trade between suppliers and firms in the countries engaged in this trade friction and foster entrepôt trade. While entrepôt trade could mitigate the negative effect of international trade friction, its capacity for mitigation is limited. Furthermore, international trade friction would also facilitate enterprise internalization. However, the response strategies integrating entrepôt trade, supplier diversity with in-house production also cannot absolutely counteract the adverse effect to sustain its trade volume.

53 - Federated Multiple Tensor-on-Tensor Regression for Multimodal Data under Data-Sharing Constraints

Zihan Zhang, ISyE Georgia Tech, Atlanta, GA, United States, Shancong Mou, Mostafa Reisi, Massimo Pacella, Jianjun Shi

In recent years, diversified measurements reflect the system dynamics from a more comprehensive perspective in system modeling and analysis, such as scalars, waveform signals, images, and structured point clouds. To handle such multimodal structured high-dimensional (SHD) data, combining a large amount of data from multiple sites is necessary (i) to reduce the inherent population bias from a single site and (ii) to increase the model accuracy. However, impeded by data management policies and storage costs, data could not be easily shared or directly exchanged among different sites. Instead of simplifying or facilitating the data query process, we propose a federated multiple tensor-on-tensor regression (FedMTOT) framework to train the individual system model locally using (i) its own data and (ii) data features (not data itself) from other sites. Specifically, federated computation is executed based on alternating direction method of multipliers (ADMM) to satisfy data-sharing requirements, while the individual model at each site can still benefit from feature knowledge from other sites to improve its own model accuracy. Finally, two simulations and two case studies validate the superiority of the proposed FedMTOT framework.

54 - Dynamics of Sentiment Involvement on Social Media in Emergencies: Case Studies Based on Microblog Data in China

boyu Zhu, Shanghai Maritime University, Shanghai, China, People's Republic of

This study focuses on the polarization of online public opinion and its formation mechanism in the era of social media in China, especially the role of official media and we-media opinion leaders in public opinion guidance. By using case analysis method and collecting data from Sina Weibo platform, this study analyzed the polarization characteristics of public opinion in network circles of eight hot social events at different stages of public opinion development, as well as the dynamic characteristics of Chinese social media users' emotions on hot issues, and reached a conclusion on the emotional tendency of public opinion.

55 - On the Use of Clustering-Based Pseudo-Class Labels for Out-of-Distribution Detection

Yujin Lee, Sungkyunkwan University, Suwon, Korea, Republic of, Seokho Kang

Out-of-distribution (OOD) detection is the task of identifying instances that deviate significantly from the in-distribution (ID) data on which a classifier has been trained. This is crucial in practice, especially when deploying classifiers in safety-critical applications such as medical

diagnosis and autonomous driving. OOD detection methods require a labeled training dataset, where each instance is annotated with its respective class label, to train a classifier. However, acquiring these class labels is often laborious and costly in real-world situations. Instead of manually obtaining actual class labels, we can derive clusters of semantically similar instances from an unlabeled training dataset and use them as pseudo-class labels. In this study, we investigate the effectiveness of various OOD detection methods when pseudo-class labels are used to annotate the training dataset. We compare the suitability of different clustering algorithms with varying numbers of clusters for generating pseudo-class labels, evaluating their effect on the performance of OOD detection methods. We conduct experiments using OpenOOD benchmarks.

56 - Optimal Production, Fuel Economy Investment and Credit Trading Decisions Under Dual-Credit Policy

Zhenxiao Wang, Tsinghua-Berkeley Shenzhen Institute, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China, People's Republic of, Fen Xu, Li Xiao, Peng Yang

We study a production system that produces fuel vehicles and new energy vehicles during one selling season under the dual-credit policy (DCP), which is issued to reduce the greenhouse emissions generated from production. One specific feature of our model is that production generates two credits, either positive or negative, and DCP requires nonnegative credits at the end of the selling season and penalizes the negative credits. We show that the optimal production, fuel economy investment, and credit trading decisions depend on the initial capability of generating credits. Based on the optimal decisions, we find that DCP promotes the production of automakers who sell credits and reduces the production of automakers who purchase credits. We next discuss various factors affecting the optimal decisions and the effectiveness of DCP. In particular, a high initial capability to generate positive credits induces a \textcolor{blue}{ high profit} and large production quantity. A high credit purchasing price induces a \textcolor{blue}{low profit} and a small credit trading volume. However, the optimal fuel economy decision and the credit trading decision are non-monotone in the initial capability to generate credits, and the production quantity and the fuel economy investment are non-monotone in the credit purchasing price.

57 - Explainable AI for Risk Evaluation: Improving Projection Accuracy of Film Revenue

Doyinsolami Osobukola, The University of Tulsa, Tulsa, OK, United States, Ismail Abdulrashid

Considering the evolving landscape of the U.S. film industry, characterized by shifting consumer preferences towards streaming platforms and a projected 7.2% annual growth rate through 2030, the importance of accurate predictive analytics cannot be understated for stakeholders seeking to optimize profitability in traditional theater releases. This poster presentation employs advanced machine learning techniques including Artificial Neural Networks, Gradient Boosting, and Support Vector Machines to forecast financial risks associated with new movie releases, utilizing a dataset comprising 129 North American films released between 2016 and 2022. The results indicate that Gradient Boosting outperforms other models significantly in terms of accuracy, precision, F1 score, and area under the curve. Furthermore, the application of an explainable analytics framework, specifically Shapley Additive Explanations (SHAP), enhances the transparency and interpretability of the influence of various factors such as genre, budget, and online media promotional strategies on pre-release movie performance. This study highlights the efficacy of advanced predictive models and illustrates the potential for smarter decision-making in film production and marketing strategies, laying the groundwork for optimizing financial outcomes in the film industry.

58 - Optimal Timing for Installing a Remanufacturing and Recycling Facility Amid Sourcing Price Uncertainties and Product Lifecycle Considerations

Mohammad Ahnaf Sadat, Iowa State University, Ames, IA, United States

In this study, we investigate the strategic decision-making process for installing a joint remanufacturing and recycling facility, focusing on turbine blade production for gas turbines as a specific example. We analyze the uncertainties in sourcing costs, which include purchasing raw materials and acquiring used products from customers, treating these as dynamic stochastic variables. In contrast, the product lifecycle and the quality of acquired used products are considered dynamic and deterministic. We employ a real options analysis, specifically utilizing Monte Carlo simulation methodologies, to model and derive solutions. The objectives of this study are to determine the optimal conditions for installing such facilities, evaluate the strategic value of this flexibility, and explore how various model variables influence these decisions.

59 - Option Valuation of Energy Storage Integration: A Real Options Approach

Gazi Nazia Nur, Iowa State University, Ames, IA, United States

This poster evaluates the economic viability of integrating energy storage systems into wind energy operations, given the inherent unpredictability of wind power that often leads to mismatches between energy supply and demand. Utilizing a real options approach, we assess the financial benefit of energy storage, taking into account uncertainties in electricity demand, wind speed, and capital costs. Our focus is on the potential for energy storage to enhance the profitability of wind power through energy arbitrage and frequency control. We employ a Monte Carlo simulation and use the least square method to estimate the option value of energy storage. The goal is to determine if the additional benefits can justify the investment in energy storage, thereby providing insights into its financial feasibility within the wind energy sector.

60 - A Deep Learning Framework for High-Dimensional Time Series Forecasting

Ali Sarabi, Arizona State University, Tempe, AZ, United States, Arash Sarabi, George Runger

Accurate long-term time series forecasting remains a challenging task in various real-world applications, requiring the consideration of high-dimensional feature variables. Traditional forecasting methods often rely on manual feature selection, which can be time-consuming and subjective and may not capture the intricate relationships among variables. While advanced deep learning methods have significantly enhanced our understanding of complex data dynamics, it is important to note that time series possess distinct properties such as trend, seasonality, and autocorrelation, which differentiate them from other data sequences. These specific characteristics should be carefully taken into account in the modeling process. To address these challenges, we propose a novel architecture that combines a graph neural network block with a \ac{LSTM} encoder-decoder block for end-to-end feature learning and forecasting in multivariate time series forecasting. This architecture learns complex data relationships while considering time series properties, resulting in accurate and reliable predictions. Our evaluation of this framework on ETDataset benchmark datasets for sequence forecasting tasks shows forecasting accuracy and reliability comparable to prior state-of-the-art models. This framework helps decision-makers make informed decisions with deep learning models based on the best practices in time series analysis.

61 - Post-disaster Restoration Problem of Water Distribution Systems Under Incomplete Information

Peihong Xiao, National University of Singapore, Singapore, Singapore

Restoring water distribution systems is one of the most critical and complex tasks in post-disaster recovery. Water pipe restoration is essential for ensuring the supply of clean water to affected areas, which is crucial for relief distribution and public health. Given the limited resources, uncertainty, and urgency during disaster response, efficient and effective planning for water pipe restoration is vital to achieve connectivity between water supply and demand. In this paper, we define the stochastic water pipe restoration problem (SWPRP), which addresses post-disaster scenarios where limited information on water pipe damage is updated as restoration progresses. The primary decision in SWPRP is determining the sequence of pipes to restore in each period to maximize the benefit of satisfying water demand. To optimally solve SWPRP, we develop a partially observable Markov decision process (POMDP) model, derive several structural results, and propose a heuristic to reduce computational complexity. We compare our approach with three benchmark recovery strategies (static-importance based, damage-based, and distance-based) and test the performance on randomly generated instances reflecting various geographical and information settings. The results demonstrate the effectiveness of our method in reducing post-disaster losses.

62 - Fair and Efficient Chores Allocation via Earning Restricted Equilibria

Aniket Murhekar, University of Illinois, Urbana-Champaign, Champaign, IL, United States

We study the problem of allocating a set of indivisible chores among n agents in a fair and efficient manner. The agents have additive valuations. We consider commonly sought-after envy-based fairness notions and the efficiency notion of Pareto-optimality (PO). For the strongest relaxation of envy-freeness, namely envy-free up to any item (EFX), no better than O(n^2)-approximation is known, and achieving a constant approximation is a major open question. For the slightly weaker relaxation, namely envy-free up to one item (EF1), proving the existence of an allocation that is simultaneuously EF1 and PO allocation is another major open question; in fact the existence of \alpha-EFk and PO allocations is unknown for any constants \alpha and k! Extensive work on these problems have tackled special cases, like n=3 agents or bivalued instances, by leveraging on their special structures.

In this work, we make significant progress on both problems: for all chore allocation instances, we show the existence of (i) 5-EFX allocations, and (ii) 2-EF2 and PO allocations. Technically, these algorithms are built upon on the novel concept of earning-restricted (ER) competitive equilibrium, which limits agents' earnings from each chore. Showing the existence of an ER equilibrium turns out to be quite non-trivial, for which we derive a linear complementarity problem (LCP) formulation and show that the classic complementary pivot algorithm on the LCP terminates at an ER equilibrium. This gives the first-ever constant-factor approximation for EFX and the first non-trivial approximation for EFk+PO, that too a constant one, for constant k.

63 - A Bilevel Model to Configuring Rail-truck Intermodal Network for Dangerous Goods Shipments

Nishit Bhavsar, McMaster University, Hamilton, ON, Canada, Elkafi Hassini, Manish Verma

Rail and truck serve as the main transportation methods in North America. However, employing them as separate modes presents problems such as traffic congestion from widespread truck use and limited rail access due to infrequent rail lines. Combining these modes has the ability to overcome challenges and limitations in individual modes. As a result, intermodal rail and truck has emerged as a complement to individual modes. A key to the efficient operation of rail-truck intermodal transportation is determining associated activities such as inbound drayage, long-haul, and outbound drayage. In this research, our focus is primarily to: i) identify terminals for rail haul operation from the perspective of cost, and ii) determine routings for inbound and outbound drayage activities from the perspective of risk. To achieve this, there are a few studies that design intermodal networks as a hub and spoke system for regular freight. We extend the concept of intermodal hub and spoke networks to tactical planning of dangerous goods shipments. To this end, we develop a bilevel model in which the upper level is a *p*-hub median single allocation problem to minimise the total transportation cost and the lower level is a routing problem for intermodal shipments to minimise the total transportation risk in the network. We reduce the problem to a single level and illustrate the solution on a prototype intermodal network. We include a case study to demonstrate its application in one of the key goods movement corridors in Canada.

64 - Bayesian Distributionally Robust Optimal Investment Banking Contract Design

Ziheng Su, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Huifu Xu

hello

65 - A Tri-Mode Delivery System to Support Reliable and Efficient Last-Mile Deliveries

Minghao Chen, Columbia University, New York, NY, United States

In the rapidly evolving landscape of logistics and transportation, innovative delivery systems are becoming essential to meet the increasing demand for efficient, timely, and flexible solutions. One such cutting-edge approach is the integration of truck-drone-robot delivery systems. This hybrid model leverages the strengths of each component—trucks, drones, and robots—to create a seamless and highly efficient delivery network.

Trucks serve as the backbone of the system, providing long-haul transportation and acting as mobile distribution hubs. They carry large volumes of goods over extensive distances and can position themselves strategically within urban and suburban areas.

Drones bring agility and speed to the system, capable of bypassing traffic and reaching destinations quickly, particularly in densely populated or hard-to-reach areas. They are ideal for delivering lightweight packages over short distances, directly from the truck to the customer's location.

Robots, particularly ground-based autonomous robots, add precision and reliability to the last-mile delivery process. These robots can navigate sidewalks and streets, delivering packages right to the doorstep. They are equipped with advanced sensors and navigation systems, ensuring safe and accurate delivery even in complex urban environments.

By integrating these three components, the truck-drone-robot delivery system offers a versatile and scalable solution that enhances delivery efficiency, reduces costs, and improves customer satisfaction. This synergy of technologies represents a significant advancement in the field of logistics, paving the way for the future of urban and suburban delivery services.

66 - Modeling Electric-vehicle Infrastructure Resilience Using an Agent-based Simulation Approach

Adelyn Heuer, Wichita State University, Wichita, KS, United States, Ehsan Salari

In recent years, there has been a significant increase in the sale of electric vehicles (EV), surpassing a market share of 6% for new vehicle sales in the US. The EV transition is accelerating at a time of growing concerns for climate change. The frequency and severity of natural disasters and extreme weather events have intensified over the past decade and the trend is predicted to continue. Extreme weather events often lead to sudden spikes in charging demand and widespread power outages, potentially disrupting power supply to households and charging stations. The goal of this research is to investigate the resilience of the EV charging infrastructure in the State of Kansas.

We employ agent-based simulation modeling to characterize the uncertainty associated with the decision-making process of individual EV owners during both normal and extreme conditions and use the model to study emergent patterns in the EV charging demand and the potential strain put on the EV charging infrastructure in the State of Kansas. The simulation model emulates major Kansas highways and interstates, establishing a network of cities and DC Fast charging stations to represent potential routes and stations available for statewide transportation.

The simulation model is used to identify any deficiencies in the current EV charging infrastructure in the State of Kansas and assess the performance of resilience-building interventions, including the design and operation of fixed and mobile charging stations as well as planning public information campaigns for managing EV-charging demand during disasters.

67 - An Exact and Efficient Algorithm for Basis Pursuit Denoising via Differential Inclusions

Gabriel Provencher Langlois, New York University, New York City, NY, United States, Jerome Darbon

Basis Pursuit Denoising (BPDN) is a cornerstone of compressive sensing, statistics and machine learning. Its applicability to high-dimensional signal reconstruction, feature selection, and regression problems has motivated much research and effort to develop algorithms for performing BPDN effectively, yielding state-of-the-art algorithms via first-order optimization, coordinate descent, or homotopy methods. Recent work, however, has questioned the stability, accuracy and efficiency of these state-of-the-art algorithms for BPDN. For example, the glmnet package for BPDN, which is state-of-the-art due to its claimed efficiency, suffers from instability and can yield inaccurate solutions that lead to many so-called false discoveries. Another example is existing homotopy methods for BPDN; most require technical assumptions that may not hold in practice to compute exact solution paths. Without a stable, exact, and fast algorithm, these shortcomings will continue to hinder BPDN for high-dimensional applications. Here, I will present a novel homotopy algorithm based on differential inclusions that is stable and performs BPDN exactly (numerically, up to machine precision). We will present some numerical experiments to illustrate the efficiency of our algorithm and discuss various theoretical implications of our algorithm.

68 - Net-Zero EV Charging Infrastructure for Texas through 2040

Valentina Mora, Texas State Univerisity , San Marcos, TX, United States, Tongdan Jin

The "Sizing Renewable Microgrid for Net-Zero and Resilient EV Charging of Texas through 2040" project aims to address the pressing need for a sustainable and reliable electric vehicle (EV) charging infrastructure in Texas. With the Electric Power Research Institute (EPRI) projecting that EVs could constitute up to 55 percent of all vehicles in Texas by 2040, there is an urgent demand for a forward-looking solution. Furthermore, this growth in EV adoption presents the opportunity to significantly reduce carbon emissions if these vehicles are charged using renewable energy sources.

This project's primary objectives are to plan and design a net-zero statewide EV charging infrastructure, implement reliable and cost-effective charging stations, and integrate wind turbines and solar photovoltaic (PV) microgrid systems to ensure power resilience and zero carbon emissions. By achieving these goals, we will contribute to Texas' sustainability, reduce carbon emissions, and create a blueprint for future EV charging infrastructure development.

69 - Optimizing Policy Decisions in Interconnected Populations: A Spatially-Dependent Framework for Dynamic Decision-Making

Jimi Kim, University of Texas at Dallas, Dallas, TX, United States, Yunan Wu

Dynamic decision-making is crucial in policy formation. In reality, states or countries typically make these decisions, and their effects are interdependent due to spatial relationships. However, conventional policy decision-making strategies often overlook these spatial effects, assuming no interference among individuals. In this project, we propose a novel framework to measure these interdependent effects across multiple stages. We also propose the methods to estimate the optimal policy regime taking the spatial effects in consideration for a series of time spots, as well as constructing confidence intervals for the coefficients indexing the optimal policy regime based on multiplier Bootstrap methods. We rigorously demonstrate the validity of the proposed estimation and inference methods.

We apply the proposed method on a case study on COVID-19 school closing policies. We develop a composite outcome variable that balances public health concerns with economic activity. This outcome metric combines the Standardized Incidence Ratio (SIR) for COVID-19 cases with a mobility index, facilitating precise policy interventions by each state aimed at reducing confirmed cases without excessively restricting mobility. By applying our methodologies to real-world data on state-level interventions and outcomes, we illustrate how this approach can optimize policy decisions. Our findings have substantial implications for enhancing the effectiveness of personalized intervention strategies in interconnected populations. This approach has the potential to revolutionize precision policymaking in spatially dependent contexts while maintaining a balance between public health and economic considerations.

70 - Meta-Analytic Predictive Priors for 2 X 2 Tables with Structural Zeros

Will Stamey, University of Notre Dame, South Bend, IN, United States, James Stamey

High quality prior information can lead to better decision making by improving precision in estimation and by providing realistic estimates to power studies. This is especially critical in business and healthcare settings where trials are expensive or risky. One approach to using historical data to develop prior distributions is the meta-analytic predictive prior. We propose a hierarchical Bayesian model for the focal parameters in a 2x2 table with a structural zero. This is a commonly used model for correlated proportions such as in treatment sequences where only non-responding (or only responding) subjects are treated a second time. We investigate the performance of the hierarchical model via simulation. We then illustrate the usefulness of the model by showing how a set of historical studies can be used to build a predictive distribution for a new study that can be used as a prior distribution for both the rate ratio and marginal probability of a positive test.

71 - A Dark Side of Trust: Examining the Influence of Environmental Risk Perception on Citizens Plastic-Avoiding Behavior

Bairong Wang, Shanghai Maritime University, Shanghai, China, People's Republic of, Bin Liu, Yong Li

This study explores the influencing dynamics of environmental risk perception on plastic-avoiding behavior by incorporating government trust and environmental locus of control within the influencing structure. Via an online survey, this study received 1126 valid responses and used partial least square-based structural equation modeling (PLS-SEM) techniques for data analysis. Three major findings are obtained. First, environmental risk perception positively impacts people's plastic avoiding behavior. Second, this relationship is partially mediated by environmental locus of control. Finally, government trust moderates the impact of environmental risk perception on both the environmental locus of control and plastic-avoiding behavior. When government trust is higher (lower), environmental risk perception has less (more) influence on the environmental locus of control and plastic-avoiding behavior. Therefore, absolute high government trust is far from ideal in environmental management because it induces high government dependence, which stimulates people's "inertia" and makes them shirk their responsibilities for environmental protection. To reduce the dark side of government trust, it is suggested that the government shows some "weakness" and emphasizes its need for the public's support for plastic crisis management.

72 - Massively Parallel Environments for Large-scale Combinatorial Optimizations Using Reinforcement Learning

Deron Diaz, Rensselaer Polytechnic Institute, Troy, NY, United States, Joshua Pile, Philip Paterson, Ethan McCartney, Xiao-Yang Liu, Ming Zhu

Combinatorial Optimization (CO) problems--such as finding the MaxCut of a graph, ground state of the Ising model, or optimal circuit designs--are NP-hard, making high-quality solutions challenging due to the exponential growth of the search space. Since quantum circuit and digital circuit optimization are theoretically equivalent, finding optimal quantum circuits--a crucial problem in quantum computing--also falls into this category. Traditional methods--heuristics and evolutionary algorithms--are often inefficient. Reinforcement Learning (RL) offers a promising alternative, utilizing powerful search capabilities and GPU-accelerated training. However, CPU-based RL environments face significant sampling efficiency bottlenecks.

This study applies a GPU-based sampling method to address the scalability issues of CPU-based sampling. The parallel processing power of GPUs causes the speed for obtaining high-quality samples to increase significantly. This is crucial for large and complex graphs--like the MaxCut problem--and other NP-hard problems like quantum circuit optimization, where processing different vertices and edges in parallel is markedly faster than CPU-based methods.

The results of the study highlight an empirical speedup of three orders of magnitude in the sampling process compared to CPU-based methods. It is demonstrated that GPU-based sampling greatly enhances the efficiency of finding high-quality solutions for large CO problems. Our implementations support many parallel environments--ranging from 8,000 to 16,000--enabling a 100x speedup in the training process. Key contributions include the development of GPU-based massively parallel environments for MaxCut; and the implementation of MCMC simulations on GPUs through the use of GPU tensors for sampling to eliminate the CPU-GPU communication bottleneck.

73 - Unsupervised Promptable Defect Segmentation In Laser Additive Manufacturing

Israt Zarin Era, West Virginia University , Morgantown, WV, United States, Imtiaz Ahmed, Srinjoy Das , Zhichao Liu

In the domain of Laser Additive Manufacturing (LAM), precise image-based defect segmentation is crucial for ensuring product quality and enabling real-time process control. However, this task presents several challenges, including the absence of labels and the overhead of supervised fine-tuning of models. To address these challenges, we've devised a novel framework for image segmentation employing a state-of-the-art Vision Transformer (ViT)-based Foundation model, known as the Segment Anything Model. Foundation models are currently driving a paradigm shift in computer vision tasks for various fields including biology, astronomy, and robotics among others, leveraging user-generated prompts to enhance their performance. Our approach incorporates a novel multi-point prompt generation scheme through unsupervised clustering eliminating the need for data labels or supervised model fine-tuning. We apply this framework to porosity segmentation in a case study of laser-based powder bed fusion (PBF) and achieve high Dice Similarity Coefficients (DSC) without the requirement for supervised fine-tuning in the model.

74 - Logistics of Infrastructure Network Restoration under Incomplete Information

Subhojit Biswas, Texas A&M University, College Station, TX, United States, Bahar Cavdar, Joseph Geunes

In this paper we study repair crew routing problem for infrastructure network restoration under incomplete information. We consider a disrupted infrastructure network where the dispatchers know the locations of service losses but not necessarily the actual fault locations. We model this problem as a finite horizon Markov decision process and develop efficient solution methods using reinforcement learning methods. Our solution method uses structural results and state-space aggregation to address the time-sensitive nature of such decisions.

75 - A Graph-Theoretic Approach to Scenario Reduction for Stochastic Generation Expansion Problems

Sara Shaddel, Pennsylvania State University, State College, PA, United States, Seth Blumsack

Operations research and optimization play critical roles across numerous fields, from energy systems to financial portfolio management. In stochastic programming, the challenge of managing large sets of scenarios to represent uncertainties significantly complicates decision-making processes. This research introduces a novel graph-theoretic approach to scenario reduction aimed at streamlining these complex models, thus enhancing the robustness and efficiency of decision-making under uncertainty. We implemented this method in a electricity generation expansion planning (GEP) problem, demonstrating its effectiveness in optimizing strategic decisions in energy systems.

76 - Modeling and Design of Net-Zero EV Charging Infrastructure through 2040: A Case Study of Texas

Valentina Mora Casallas, Texas State University, San Marcos, TX, United States, Saha Joy, Arslan Abbasi, Tongdan Jin

This research project aims to address the pressing need for a sustainable and reliable electric vehicle (EV) charging infrastructure in Texas. According to the projection made by Electric Power Research Institute (EPRI), EV could constitute up to 55 percent of all vehicles in Texas by 2040. There is an urgent demand for a forward-looking solution for electric vehicle service equipment planning and deployment. Furthermore, the fast growth in EV adoption presents the opportunity to significantly reduce carbon emissions if the vehicle fleet is charged using renewable energy sources, such as wind and solar energy. This project's primary objectives are to plan and design a net-zero statewide EV charging infrastructure, implement reliable and cost-effective charging stations, and integrate wind turbines and solar photovoltaic (PV) microgrid generation to ensure power resilience and zero carbon emissions. Based on the projected EV fleet size, we use the EVI-Projection Tool to determine the statewide Level 2 and Level 3 chargers. Then, the sizing and siting of solar and wind capacity to power the entire infrastructure is obtained. By achieving these goals, we will contribute to Texas' sustainability, reduce carbon emissions, and create a blueprint for future EV charging infrastructure development.

77 - Assessing the Feasibility and Economics of Seasonal Hydrogen Storage in the Power Market

Yuheng Zhang, Peking University, Beijing, China, People's Republic of, Guannan He

Hydrogen has increasingly been an attractive energy vector in the context of carbon neutrality. With the fast development of relative technologies in the hydrogen supply chain, including generation, transportation, storage, and distribution, the hydrogen economy, of which the trade scale is estimated to be 18 EJ (LHV) in 2050 according to IRENA, is emerging in the global range. In particular, in the power system, hydrogen storage has long been considered a solution to long-term and short-term storage to balance intermittent renewable generation and uncertain demand. However, the economics of hydrogen storage have not been systematically revealed, jeopardizing the deployment of hydrogen. In this paper, we deeply dive into hydrogen's economic viability in the power markets using a general energy arbitrage optimization model. Three different market schemes including day-ahead, time-of-use, and deduced seasonal price schemes are compared with different hydrogen storage configurations. Underground caverns, gaseous tanks, and liquid hydrogen containers are considered to reveal the difference in economics for different hydrogen storage types. The results reveal that the seasonal price is appropriate for prompting hydrogen deployment, especially in long-term energy storage applications. The ratio of revenue to costs could reach about 0.8 for gas hydrogen storage and more than 1 for underground cavern and liquid hydrogen storage in our derived seasonal price scheme, meaning that these types of hydrogen storage are profitable for investors with a proper market scheme.

78 - Learning Real-Time Intention from Noisy Behavioral Data via State Space Model

jiaxuan jiang, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Jiapeng Liu, Milosz Kadzinski, xiuwu Liao

Dynamic intention recognition has been utilized across various fields, such as autonomous driving, e-commerce, human-machine interaction, and military strategy, to understand the evolving intentions. It involves analyzing the agent's behavior over time, which poses challenges in mapping behaviors to intentions, handling noisy continuous data, and interpreting intentions in real-time. To overcome these obstacles, we introduce a novel intention recognition approach using a state space model. Our method visualizes the generation process with a directed acyclic graph, providing clear insights into dependencies and transitions. We employ Kalman filtering to model the evolution of behavior and filter out noise. Additionally, we create a classification system for predicting intentions. Extensive experiments with real and synthetic datasets demonstrate our model's effectiveness, especially in handling imbalanced data and minority classes, making it a practical solution for real-time intention analysis of moving agents.

79 - Exploring Nash Equilibrium in Airline Frequency Competition through Second-Order Cone Programming and Code-Share Agreements

Wen-Fang Wang, National Tsing Hua University, Hsinchu, Taiwan, Yu-Ching Lee, Chun-Han Wang

In the contemporary aviation industry, strategically determining flight frequencies between airports is a crucial factor that impacts market share, operational costs, and the competitive dynamics among airlines. The rapidly evolving aviation market has transformed commercial relationships from mere competition to the formation of alliances and collaborations. Among these cooperative strategies, the code-sharing agreement stands out as particularly significant.

Code-sharing enables airline to sell seats on flights operated by partner airlines, thereby enhancing participation through commercial agreements. This practice is commonly used at hub airports to enhance connectivity between long-haul and high-frequency short-haul flights. Statistics reveal that over half of the airlines in U.S. market have adopted code-sharing strategies, particularly on routes near major hub airports.

Despite the widespread use of code-sharing in the airline industry, there is a notable absence of a comprehensive quantitative model that simultaneously addresses both codeshare strategies and flight frequency competition. Recognizing this gap, this study aims to develop a frequency competition model that integrates codeshare strategies within a game-theoretic framework. The model utilizes second-order cones to represent certain constraints. To determine the Nash Equilibrium for this problem, we leverage the optimality conditions of second-order cone programming to construct a KKT model that includes primal, dual, and complementary slackness constraints. The proposed model performs quantitative calculations to determine the optimal flight frequency for each route segment and offers strategic insights regarding codeshare agreements, selecting suitable partners, and profit sharing among airlines.

80 - A Nash Equilibrium Algorithm Using Rational Generating Function to Solve Congestion Games

SHAN YUN-CHEN, National Tsing Hua University, Hsinchu, Taiwan, LEE YU-CHING, WANG CHUN-HAN

Game theory finds wide application in economics, computer science, and engineering, especially in analyzing participants' competitive or cooperative strategies. While previous research has proposed various methods for different games, an effective approach to finding all Nash equilibria in integer programming (IP) games remains elusive.

In this study, we implemented a novel algorithm capable of enumerating all pure Nash equilibria on a polyhedron. Building on existing research, when the number of players is fixed, it becomes feasible to derive a generating function containing all pure Nash equilibria and enumerate all equilibrium points in polynomial time using binary search and residue techniques. Our algorithm integrates Python, SageMath, and LattE for implementation.

To demonstrate algorithm performance, we use congestion games as an example, which are IP games based on network congestion problems with applications in traffic management, network traffic control, and supply chain management. Experimental results demonstrate that our algorithm accurately and swiftly identifies all Nash equilibria, and outperforming brute-force enumeration in solving speed. This advancement not only enhances solution methodologies for Nash equilibria in congestion games but also offers effective solutions for domains like traffic management and network optimization.

81 - Fair Allocation of Indivisible Chores: Beyond Additive Costs

Li Bo, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

We study the maximin share (MMS) fair allocation of m indivisible chores to n agents who have costs for completing the assigned chores. It is known that exact MMS fairness cannot be guaranteed, and so far the best-known approximation for additive cost functions is $\frac{13}{11}$ by Huang and Segal-Halevi; however, beyond additivity, very little is known. In this work, we first prove that no algorithm can ensure better than $\min\{n,\frac{\log m}{\log \log m}\}$-approximation if the cost functions are submodular. This result also shows a sharp contrast with the allocation of goods where constant approximations exist as shown by Barman and Krishnamurthy and Ghodsi et al. We then prove that for subadditive costs, there always exists an allocation that is $\min\{n,\lceil\log m\rceil\}$-approximation, and thus the approximation ratio is asymptotically tight. Due to these hardness results for general subadditive costs, we turn to study three specific subadditive costs, namely, vertex-covering, bin-packing and job-scheduling. We show that constant approximate allocations exist for all these settings.

82 - Proportional Response in Stochastic Online Exchange Economies

Chia-Ying Lin, National Tsing Hua University, Hsinchu, Taiwan, Yu-Ching Lee

Problem Definition: This study focused on periodic exchange economies, referred to as Arrow-Debreu market, where each player acts as both a seller of a divisible good and a buyer with a certain amount of money. The value of goods fluctuates stochastically over time. Each day, players bring their goods and submit bids on the goods they prefer to maximize their utilities. Goods are allocated based on the bids provided by the players, and each seller collects the bids on their goods, which becomes their budget for the next period.

Methodology: we introduce a distributed auction algorithm, termed the “lazy proportional response”. This algorithm provides a more general framework for proportional response dynamics. It allows each player to update their bids based on the contribution of each good to their utility. Additionally, the framework requires players to save a fractional amount of money for later use.

Managerial implications: The lazy proportional response algorithm addresses privacy concerns by allowing buyers to adjust their bids directly, without revealing sensitive information. Moreover, it resolves the cyclic issues associated with linear utilities, leading to convergence in everything: utility, consumption, and prices.
Additionally, our algorithm demonstrates robust performance guarantees across various scenarios, making it universally applicable to diverse environments.
Key words: exchange markets, proportional response, online, market equilibrium

83 - The Cost of Supply Chain Complexity: Quantifying the Effect of Product Variety on Profit Margin

Daniela Ruiz, Pennsylvania State University, University Park, PA, United States

What is the cost of complexity in the supply chain? When companies assemble products to order, the products can potentially be highly customized to fit a specific clients’ needs. This strategy leads to a high number of possible product configurations, adding complexity to decisions along the entire supply chain. This research focuses on how supply chain complexity affects costs and therefore profit margin. This problem is exemplified and an approach for its analysis is proposed through a research project in collaboration with senior management and directors from IBM’s storage division. Through a two-stage stochastic linear program, demand and manufacturing data are used to model different possible demand scenarios and find optimal decisions for production with product conversion and expedited manufacturing used as recourse decisions once demand becomes certain. The resulting cost and revenue are compared for different demand outcomes changing the number of products in the offering. Comparison is to be held between scenarios which offer a highly customizedproduct portfolio and ones with a higher level of standardization and fewer products.

84 - Optimizing Border Security Strategies: A Game-Theoretic Framework

Yusuf Ihsan Tokel, University at Buffalo, Buffalo, NY, United States, Jun Zhuang, Yiqi Zhao

The United States of America has been the top country for asylum seekers for decades. Although the great majority of the origin countries are still from countries in America continents, recent trends show that more people all over the world after the COVID lockdown seek their chances to seek asylum from the United States. In this work, we investigated the most recent trends using United States Customs and Border Protection (CBP) data and generated insights to grab the attention of the field professionals and the general public to this matter. There are plenty of studies to focus on the trends from Mexico or the Northern Triangle of Central American countries. Still, to the best of our knowledge, a holistic effort to investigate the trends from the countries outside of the American continent has not been put into place. With this work, we aim to fill this gap by analyzing potential routes from various origin countries to South-Central and North America, assessing the risk factors for each pathway, and modeling the interaction between unauthorized immigrants/smugglers and the government using decision and game-theoretic framework

85 - Data-driven Approach in Hub Network Design with Demand Uncertainty

Davod Hosseini, Sobey School of Business, Saint Mary's University, Halifax, NS, Canada, Fahimeh Rahimi, Borzou Rostami, Saleh Farham

This study addresses a vehicle-based hub network design problem under demand uncertainty, with primary applications in freight distribution and parcel delivery systems. The focus is on leveraging economies of scale through effective vehicle utilization for consolidated freight movement. Key aspects include selecting hub nodes, assigning demand nodes to these hubs, determining the type and number of vehicles between hubs, and managing vehicle allocation to demand nodes to minimize total costs. Considering the impact of customer demand on vehicle utilization, the study incorporates demand uncertainty into the planning phase to ensure coverage.
To tackle the inherent complexity of demand uncertainty, we employ a data-driven strategy. This approach approximates stochastic expected costs based on historical demand data, utilizing machine learning techniques such as k-nearest neighbors for improved accuracy. Our methodology ensures that the optimal solution of the data-driven approximation closely aligns with the underlying optimization problem. However, the approximation becomes challenging with an increasing number of demand points or samples. To mitigate this, we introduce a Benders decomposition approach with efficiently solvable subproblems, enhancing the tractability of the solution process. This poster highlights our innovative methods and demonstrates their effectiveness in addressing the complexities of the hub network design problem under demand uncertainty.

86 - Optimizing Procurement: An Exact Approach to Commodity Classification with Strategic Criteria

Frederik Vos, Portland State University, Portland, OR, United States, Bita Mirzaei, Katharina Olkis

In practice, different types of buyer-supplier relationships and commodities need to be classified & managed in different ways. Approaches like the Kraljic Portfolio Matrix (KPM) and other classification tools have been developed to guide managers to make the right choices. However, managers are often confronted with the limitations of qualitative analysis methods related to these tools, particularly when increasing variables and boundary conditions have to be considered when classifying different groups to manage.

In order to address this, we introduce the Generalized Capacitated Procurement Commodity Classification Problem (GCPCCP) with strategic criteria. We also revisit the traditional employment of Clustering algorithms for supplier segmentation within the KPM, highlighting alternatives that, unlike Clustering, are designed to optimize supplier categorization directly in alignment with strategic goals. This comparative study not only illuminates the unique contributions of each method, but also enhances decision making in supply chain management by presenting a more structured and quantitative analysis within the KPM framework.

Summarized, we aim to contribute to literature by introducing the GCPCCP with strategic criteria, also providing a discussion of these potential strategic criteria (such as sustainability targets, attractiveness / preferences, quality assurance criteria and interdependency constraints), focusing on their economic impact. We will also show the results of the first model approach, based on the Capacitated Facility Location problem, assuming heterogeneous facilities with a set of unique attributes. We also provide the first set of benchmark instances. These findings will be presented in a poster-form.

87 - A Decision Support System for Anesthesia Staffing Optimization

Annajirao Challa, Binghamton University (SUNY), Binghamton, NY, United States, Kyle Hayes, Wen Cao, Mohammad Khasawneh

Effective management of Operating Room (OR) schedules and anesthesia staffing is crucial for hospital efficiency and patient care. This project aims to develop a decision support system to evaluate OR case volumes and calculate staffing capacity and demand across hourly time intervals by day of the week (DOW). The study includes the staffing of Certified Registered Nurse Anesthetists (CRNAs), W2 per diems, 1099 contractors, and locum tenens. This project aims to align staffing levels with workload demands, aiding in better resource utilization and patient flow. The findings support the development of an optimized work schedule, facilitating coordination between surgical teams and anesthesia providers. This project enhances operational effectiveness and highlights the importance of data-driven decision-making in optimizing surgical services.

88 - Non-metrical Similarity-based Clustering: Minimal Clique Cover with Minimal Edge-cuts

Yuxuan Ren, Georgia Institute of Technology, Atlanta, GA, United States

This research addresses a non-metrical similarity-based clustering problem, focusing on maximizing within-cluster similarity and minimizing inter-cluster similarity. The primary objective is to ensure that nodes within the same cluster exhibit similarity above a specified threshold. The secondary objective is to minimize the number of clusters while maintaining the within-cluster similarity. The tertiary objective is to minimize the number of similar node pairs across different clusters, given that the within-cluster similarity and minimal number of clusters are achieved. The clustering problem is formulated as a compact Mixed Binary Integer Programming (MBIP) model with a minimal number of variables and constraints. Despite this, the MBIP remains complex, prompting a divide-and-conquer approach: First, the similarity matrix is transformed into an incidence matrix, indicating node similarity. This incidence matrix is further converted into a graph, and clustering is performed sequentially on connected subgraphs, as nodes in isolated subgraphs cannot form clusters. Second, we employ suboptimal or optimal solutions to the minimal clique cover problem, addressing the first two objectives as feasible heuristic solutions to the MBIP. This approach accelerates the identification of good feasible solutions. Last but not least, we identify the maximal independent set and assign distinct clusters to each point within this set, thereby reducing the number of variables and constraints in the MBIP.

89 - Customized Metamaterial Footwear Design Based on Plantar Stress and Finite Element Analysis

Ziyang Zhang, Oklahoma State University, Stillwater, OK, United States, Yu Feng, Chenang Liu

This work presents a data-driven approach for designing customized metamaterial-structured shoe soles, tailored to individual plantar pressure using generative models. By integrating machine learning with finite element analysis, this work is expected to identify the optimal design based on user-specific biomechanical data. Our approach generates metamaterial structures from plantar pressure, uses finite element analysis for stress-strain data, and maps relationships via machine learning. Based on plantar pressure input, generative model optimizes metamaterial thickness in different regions for target sole stiffness and cushioning, enhancing functionality, and reducing development time and costs.

90 - Explainable Failure Prediction of Hard Disk Drive Using SMART Data

Wei Li, Northeastern University, Boston, MA, United States, Haozhou Zhou, Srinivasan Radhakrishnan, Sagar Kamarthi

Timely hard disk drive (HDD) replacement is critical to prevent data loss and streamline maintenance operations in modern data centers. HDDs are equipped with Self-Monitoring, Analysis, and Reporting Technology (SMART) to track key performance indicators with empirically established operational thresholds. In recent years, various machine learning models have utilized SMART time series data for early HDD failure prediction. However, decision-makers need greater transparency and explainability to trust and implement these data-driven failure prediction models. Using SHAP (SHapley Additive exPlanations) analysis, we proposed a methodology for extracting explainable features from the SMART time series and visualizing the effect of these features on short-term HDD failure prediction. We trained an XGBoost model with top information-rich features and evaluated the failure detection rate and false alarm rate. We illustrated the effectiveness of the proposed approach on Backblaze data for Q1 and Q2 of 2023. The model provided a 74.7% failure detection rate with only a 0.73% false alarm rate on the test data from Q3 of 2023. The results demonstrate the potential of the proposed approach for developing an effective HDD failure prediction model with explainable features.

91 - A Pareto Improvement Bumping-Rescheduling Policy for Operating Room Scheduling

Hung Do, University of Vermont, Burlington, VT, United States, David Novak, Mitchell Tsai, Andrey Ukhov

For hospitals using shared operating rooms for scheduled and emergent cases, a bumping policy is needed to accommodate emergent cases as they arrive, but it often operates with competing objectives. The current bumping policies do not offer efficient access to capacity (i.e., Operating Rooms) as well as any mechanism for collaboration. We propose, model and analyze a bumping policy called First-In-First-Out (FIFO) Bump Policy in the context of Operating Room Scheduling that facilitates load balancing and more efficient capacity usage with pooling and offers a collaborative mechanism with positive behavioral effects to speedup service rates, resulting in Pareto improvement on the dimensions of interests: total tardiness, total expected overtime and total expected idle time. Our one-day implementation, numerical and simulation study show promising improvement of the proposed policy.

92 - Optimal Lock Maintenance of the U.S. Inland Waterway Transportation System under Uncertainty

Maryam Aghamohammadghasem, University of Arkansas, Fayetteville, AR, United States, Jose Azucena, Haitao Liao, Shengfan Zhang

The U.S. inland waterway transportation system (IWTS) is an essential part of the country's multimodal freight transportation network. In addition to seasonal droughts and floods, the operations of the IWTS may be disrupted by random malfunctions and scheduled maintenance of its critical components. Among these critical components, locks play a key role in the operation of navigable inland waterways, and lock-induced disruptions to the supply chains of related industries, such as agriculture and manufacturing, often result in significant economic losses. To assess the performance of the U.S. IWTS, we develop a PyNetLogo simulation tool to capture the movements and delays of cargoes under various operational uncertainties. Using this simulation tool, a lock repair and preventive maintenance strategy is determined via Deep Reinforcement Learning (DRL) to minimize the loss due to lock-induced disruptions. To illustrate the proposed modeling and decision-making method, the McClellan-Kerr Arkansas River Navigation System is considered in our case study, where a random policy and a first-come, first-served policy conventionally implemented in practice are also presented for comparison. The results show that the optimal strategy obtained by the proposed DRL-based approach outperforms the conventionally implemented alternatives in various aspects. Most importantly, the levels of availability of all the locks are significantly improved, enabling a more seamless cargo flow along these navigable inland waterways. For the benefit of stakeholders, our further study reveals that employing multiple full-time repair crews instead of one can further
increase the availability of the locks.

93 - Steeling Against Time

JeongUk SEO, Sungkyunkwan University, Suwon, Korea, Republic of, SEONG-HYUN SEO, DONG-JOON LIM

Concept drift (CD) refers to a situation in which the statistical properties of the target variable, which the model aims to predict, unpredictably change over time. Methods for detecting and adapting to concept drift provide insights into scenarios where model performance declines due to these unpredictable changes. We propose an adaptive framework that facilitates concept drift detection and adaptation, leveraging data and feature extraction, model management, hyperparameter adjustments, and artifact versioning. Additionally, a 'buffer period' is introduced to continue data collection after drift detection, during which various 'weights' are applied to update the model. An empirical case study from the Korean steel industry highlights the necessity of autonomous retraining strategies post-model deployment and the importance of considering these strategies during model development.

94 - A New Perspective on Carrier Collaboration: Collaborative Vehicle Utilization

Ilke Bakir, University of Groningen, Groningen, Netherlands, Sahand Asgharieh Ahari, Kees Jan Roodbergen

To reduce the environmental impact and costs associated with road transport, carriers can collaborate with each other in making deliveries to their customers, typically sharing any additional profits. Our proposed collaborative vehicle utilization framework allows participating carriers to borrow trucks from each other. The borrowed truck is picked up from the lender's depot and used to transport goods from the borrower's depot to the delivery locations based on optimal routing decisions. Although order sharing, where carriers can make deliveries on behalf of each other, is a common approach to carrier collaboration in the literature, it is rarely implemented in practice due to carriers' reluctance to share sensitive customer information. In contrast, our framework does not require such information exchange. We develop two integer programming formulations for the collaborative vehicle utilization problem and design branch-and-price algorithms to solve them. Profits are shared among participating carriers using a cooperative game theoretical approach, which is solved by a constraint generation algorithm. Thanks to short calculation times, our collaborative vehicle utilization framework can solve instance sizes that are more relevant to practical applications, and typically achieves profits within 5% of those obtained through order sharing.

95 - Multi Stage Genetic Algorithm Optimization for Scheduling Flexible Flow Lines with Sequence Dependent Setup Times

Aadithan Anbuvanan, Clemson University, Clemson, SC, United States, Mary Beth Kurz

This research aims to optimize scheduling in flexible flow lines using a multi-stage random keys genetic algorithm (RKGA). The primary objective is to minimize makespan. The study evaluates traditional and proposed RKGA methods, utilizing flexible flowline data to validate the hypothesis on how chromosome representation impacts the solution quality.

96 - Reducing LPs with Arithmetic-Based DMU Identification for Large-Scale DEA

QIANWEI ZHUANG, Osaka University, Osaka, Suita, Japan, Hiroshi Morita

Data Envelopment Analysis (DEA) evaluates the efficiency of units with multiple inputs and outputs using Linear Programming (LP) for benchmarking. However, this becomes computationally challenging with large datasets. This study introduces an algorithm that employs only arithmetic computations to quickly distinguish a large portion of efficient and inefficient Decision-Making Units (DMUs), significantly reducing the number of LPs required for large-scale DEA computations.

97 - Development of an AI Model for Predicting Drug ATC Code

TaeGyeong Woo, Chosun University, GwangJu, Korea, Republic of, Min Seop So, Kibet Duncan, YONGSU HAN, Hahyeon Kang, GEONU KIM, Jongho Shin

The systematic classification of drugs into the Anatomical Therapeutic Chemical (ATC) system is essential for managing drug information and ensuring the efficacy and safety of pharmacotherapies. This study presents a comprehensive computational approach leveraging advanced machine learning techniques to predict ATC codes for a broad spectrum of drug classes. Utilizing a rich dataset of molecular structures represented by SMILES strings, we processed these through the Extended Connectivity Fingerprints (ECFP) to capture the diverse chemical properties that correlate with specific ATC classifications. Our predictive models, centered on Graph Convolutional Networks (GCNs), were adeptly applied to these molecular fingerprints to forecast the ATC class labels across all categories: A (Alimentary tract and metabolism), B (Blood and blood forming organs), C (Cardiovascular system), D (Dermatologicals), G (Genito-urinary system and sex hormones), H (Systemic hormonal preparations, excluding sex hormones and insulins), J (Anti-infectives for systemic use), L (Antineoplastic and immunomodulating agents), M (Musculo-skeletal system), N (Nervous system), P (Antiparasitic products, insecticides and repellents), R (Respiratory system), S (Sensory organs), and V (Various). The methodology demonstrated robust predictive performance across these classes, highlighting the potential of integrating cheminformatics with deep learning to enhance the precision of drug classification. The results underscore the efficacy of computational models in managing extensive pharmaceutical data and suggest significant implications for drug discovery and personalized medicine. By addressing the classification challenges across all ATC categories, this interdisciplinary research sets a new benchmark for computational approaches in pharmacological sciences, offering a scalable solution for the rapid categorization of novel therapeutic compounds.

98 - Fall Detection Prediction on Keypoints Attention Based on Transformer Model

Hahyeon Kang, Chosun University, Gwangju, Korea, Republic of, KIBET DUNCAN, Min Seop So, GEONU KIM, YONGSU HAN, TaeGyeong Woo, Jongho Shin

In human activity recognition, precise and timely fall detection still emerges as a challenge especially for health concerns. This research introduces an innovative fall detection method centered on the speed of keypoints derived from human body postures using Mediapipe. Using video datasets, these keypoints are tracked continuously, enabling the calculation of their instantaneous speeds. The Transformer architecture, recognized for its prowess in sequence-related tasks, is employed with an attention mechanism that discerns critical speed pattern shifts indicative of falls. This attention design ensures adaptation capture of temporal relations among keypoints, accentuating significant moments that might suggest a fall event. Preliminary tests on established datasets show enhanced fall detection accuracy compared to conventional techniques. A standout feature of the model is its notable reduction in false positives, a recurring issue in many fall detection systems. By analyzing the speed changes of these keypoints, a comprehensive understanding of human movement dynamics and potential fall markers is achieved. Boasting a harmonious mix of sensitivity and specificity, this Transformer-based method holds considerable promise for diverse real-world scenarios.

99 - Social vs. Default Norm-Nudging: An Experimental Investigation

Kam Kamyar, The Ohio State University, Columbus, OH, United States

I study the impact of norm-nudging on value domain in an incentivized binary-lottery choice problem. A norm is traditionally defined as the effect of actions and opinions of others A Nudge is defined to be a factor that does not affect the intrinsic value (i.e., expected value) of a lottery, yet may still affect the choice (for instance, by altering attentions). In addition, the paper investigates the effect from the source of the Nudge given nudges from statistics prior subjects vs. from the experimenters when the choice problem is given. The results show that subjects tend to prefer following Social Nudges over expected value maximization and Default Nudges, shedding light on the importance of socially-informed norms in nudging applications and policy-making.

100 - Go Smart and Green: A Sustainable Traffic Management Framework Utilizing Vehicle Energy Consumption and Emission Patterns

Sudenaz Ozvural, Drexel University, Philadelphia, PA, United States, Liang Zhang

The increasing number of vehicles on roads has reduced mobility, increased energy consumption, and emphasized environmental concerns. Especially in congested urban areas, vehicles tend to consume more gasoline and emit higher levels of pollutants. In this research, we introduce a traffic signal control system that utilizes vehicle energy consumption and emission data. Our primary objective is to effectively balance and reduce vehicle energy consumption and emissions at signalized intersections while ensuring equitable allocation of resources (such as green time and "travel credits") among different vehicle types. Specifically, the system collects vehicle energy consumption and emission data through sensors installed on vehicles, which transfer this information to the control center as the input. The control center then outputs signal timing plans based on resource allocation algorithms. Additionally, fairness indices are established for various vehicle types (including trucks, regular vehicles, and electric vehicles). Each vehicle is allocated "travel credits" for future use. Simulation experiments have demonstrated that our proposed method can significantly reduce both energy consumption and emissions simultaneously.

101 - Improving Customer Demand Forecasts in Semiconductor Supply Chain

Aparna Komarla, Solidigm, Berkeley, CA, United States

Customer demand forecasts are critical to business planning, specifically for supply and operations related tasks. Due to the short-range of available data and varying market conditions, it is often challenging to accurately forecast demand across multiple Solid State Drive (SSD) product SKUs in Solidigm's (former Intel NAND business) supply chain. In this work, I propose a time-series based model that leverages vector similarity metrics, such as cosine-similarity, to identify patterns in customer demand historically. The goal is to learn the time periods that best predict customer behavior in the future and other patterns such as a demand shift from low to high density SSDs. This approach is best for short-range time-series data with a continuous dependent variable, such as Solidigm's supply chain which captures demand on a quarterly basis.

102 - Predicting Inflection Points in Semiconductor Industry Revenue Using Macroeconomic Indicators

Aparna Komarla, Intel, Berkeley, CA, United States

Semiconductor industry revenue is cyclical in nature but often hard to predict or forecast. Experts have often lamented that macroeconomic indicators such as the Gross Domestic Product or Consumer Purchasing Index, which speak to consumer confidence and market strength, inform semiconductor market inflections. Predicting such inflections is helpful for industry supply and demand planning. In this work, we explore a range of multivariate forecasting methods from a simple linear optimization model to LSTM neural networks for inflection predictions. Since research vendors such as Gartner, S&P or IDC document semiconductor indicators in a quarterly or monthly fashion, we run into the dimensionality curses since the features outnumber the timestamps available for prediction. We see that simple methods offer high predictive power and help improve decision making.

103 - Reduction of Process Takt Time for Wheels of Mining Trucks via Simulation

Junayed Pasha, Gannon University, Erie, PA, United States, Jayed Pasha, Francesca Nemetz, William Thompson, Konner McCurdy

The efficiency of production is greatly driven by process takt time and is considered while improving the efficiency of making a product. A recurring product of a number of manufacturers, such as the Wabtec Corporation, is wheels of mining trucks. The process takt time for wheels of mining trucks can vary by each cycle or batch. This study aims to identify the process takt time through simulation modeling and analysis. The study also determines where changes within the manufacturing process of wheels of mining trucks can be made to reduce the takt time. A simulation model of the operations area is developed to perform a simulation analysis, so that the product goes through the stations, and the time for completion at each station is found. A reduction in the process takt time is aimed through the simulation analysis, whether that is moving around areas or implementing different processes. If a wheel does not pass the testing station, its rework time is also taken into account. Based on the results of the simulation conducted within the study, some recommendations are outlined for resolving the long process takt time issue. The recommendations include hiring more personnel, performing more maintenance inspections or audits to catch breakdowns earlier along with defective parts, and upholding cleanliness on the shop floor.

104 - Traffic Flow Management of the State Street-Bayfront Parkway Intersection with Simulation Modeling and Analysis

Junayed Pasha, Gannon University, Erie, PA, United States, Jayed Pasha, Kyle Anderson, Alvaro Agullo San Juan

One of the common challenges faced by urban societies is traffic congestion at intersections, as it leads to wasted time, increased fuel consumption, and environmental pollution. The City of Erie in Pennsylvania faces significant congestion at the State Street-Bayfront Parkway intersection. To address this issue, this study aims to analyze and improve traffic flow at this intersection, in order to reduce congestion, travel time, and improve overall traffic fluidity. This study has developed a simulation model that replicates real-world traffic conditions at the State Street-Bayfront Parkway intersection. Several factors are incorporated in the model, such as traffic control strategies, queue lengths, traffic volume, etc. Both the simulation model and real-world data suggest that there is heavy traffic congestion northbound and moderate traffic congestion eastbound at the intersection. It is not feasible to adjust the traffic light times without increasing congestion on one street or the other. Therefore, it is recommended that a second left turn lane needs to be added for State Street. This reduction in traffic will allow for the red times to be decreased by 5-10 seconds for Bayfront Parkway, in turn decreasing congestion in both directions. It is expected that the recommendations from this study would help achieve more traffic fluidity at both State Street and Bayfront Parkway.

105 - Enhancing Breast Cancer Prediction Accuracy through Optimized Machine Learning Models and Feature Selection Techniques

Sadaf Tabatabaee, Binghamton University, Binghamton, NY, United States, Nikoo Ebrahimi, Jenan Maher Hammoudeh Albayari, Shuxia Lu

Breast cancer remains a critical global health issue and early detection with accurate prognosis is essential for improving patient outcomes. To address dataset imbalances, we employed advanced data balancing techniques and utilized Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) for optimal feature selection. We applied and compared several cutting-edge machine learning (ML) algorithms, including random forest, support vector machine (SVM), gradient boosting machine, and logistic regression.

Our comprehensive methodology involved a detailed analysis of the Wisconsin Breast Cancer Dataset (WBCD), consisting of 569 instances. We performed rigorous data preprocessing, including cleaning, normalization, and resampling, to ensure a balanced class distribution and high-quality data for model training. The learning curves for each ML model were thoroughly analyzed to assess their performance.

Among the evaluated models, the Support Vector Machine (SVM) demonstrated superior performance, achieving the highest accuracy and F1-score. It exhibited strong initial training accuracy and improved validation performance as more data was incorporated. Logistic Regression also performed well, ranking second in terms of accuracy and F1-score. In contrast, both Random Forest and XGBoost encountered overfitting issues, characterized by high training accuracy but significantly lower validation accuracy. This indicates poor generalization to unseen data and highlights the need for either additional data or the implementation of regularization techniques to mitigate overfitting.

106 - Activating Pro-environmental Norm: The Crucial Role of Social Media in Shaping Public Attention towards Environmental Advocacy

Yidi WANG, Communication University of China, Beijing, China, People's Republic of, Lin Wang

Social media platforms play a pivotal role in shaping both individuals' online behavior and offline civic conduct. In an era marked by unprecedented environmental challenges, environmental protection requires strong individual engagement and public practice. Hence, the successful implementation of government environmental policies is particularly reliant on social media to raise public awareness, promote ethics and social responsibility, and encourage participation. However, the mechanisms concerning social media serve as powerful conduits for disseminating environmental information remain unrevealed.
Our research focuses on the extent to which and how social media platforms can help enhance the effectiveness of government environmental policies in promoting individuals' pro-environmental behaviors. Based on the norm activation theory, this study examines the mediating role of residents' activated pro-environmental norms, as reflected by environmentally-themed movies online, in elucidating the relationship between government efforts and the residents' carbon emissions.
We utilize a unique dataset comprising social media interactions centered on how residents engage with environmentally-themed movies online, sourced from China's largest online community dedicated to cultural product sharing. We supplement information on government environmental efforts, residents' carbon emissions, and environmental pollution. The final dataset spans from 2006 to 2010 from 26 provinces in China. The findings reveal a significant reduction in residents' carbon emissions resulting from the government's environmental efforts, with 32.9% of this effect mediated through activating residents' pro-environmental norms.
Our study contributes to the theoretical underpinnings of environmental communication and policy application. Furthermore, the findings enhance the effectiveness of environmental advocacy by media organizations and government policymakers.

107 - Toward a Safer World: Improving Disaster Response by Identifying Disaster-Related Tweets with Advanced NLP Techniques

Sadaf Tabatabaee, Binghamton University, Binghamton, NY, United States, Mohammad Khasawneh

Twitter is essential for real-time crisis communication, but distinguishing literal from figurative language in tweets is challenging. This study develops a machine learning model to accurately identify disaster-related tweets using a dataset of 10,000 manually classified entries. The methodology encompasses several key steps to ensure the robustness and accuracy of the model. Extensive data cleaning to handle informal language specific to Twitter. Tokenization and padding are applied to standardize tweet lengths for model input. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is utilized. Model evaluation is conducted using stratified k-fold cross-validation, ensuring each fold maintains the same distribution of classes. Performance metrics include the confusion matrix, and Receiver Operating Characteristic (ROC) curve, providing a thorough assessment of the model's predictive capabilities. In this study, four models were utilized to improve the identification of disaster-related tweets. The baseline model is a simple neural network for text classification. The first tuned model (Model 2) implements a Recurrent Model LSTM. The second tuned model (Model 3) enhances the LSTM with GloVe embeddings. The third tuned model (Model 4) further improves this setup by incorporating GloVe embeddings with SMOTE data balancing and cross-validation with an extensive text cleaning approach. Among these models, Model 4, significantly enhances performance. This approach addresses vocabulary gaps and overfitting, providing reliable information for disaster response teams and improving the accuracy and trustworthiness of social media communications during emergencies.

108 - Heterogeneous Graph Sequence Neural Networks for Dynamic Traffic Assignment

Tong Liu, University of Illinois at Urbana-Champaign, URBANA, IL, United States, Hadi Meidani

Traffic assignment and traffic flow prediction provide critical insights for urban planning, traffic management, and the development of intelligent transportation systems. An efficient model for estimating traffic flow could provide a more detailed depiction of traffic dynamics across the urban traffic landscape. However, existing traffic prediction approaches are usually limited to locations where sensors are placed and cannot predict traffic flows beyond sensor locations. To alleviate this constraint, we seek to go beyond a framework that solely depends on traffic sensors, and exploit the relationship between traffic flow and the origin-destination (OD) travel demand, guided by the concept of dynamic user equilibrium. In such "physics-based" approach, we propose a novel framework, namely a Heterogeneous Spatio-Temporal Graph Sequence Network (HSTGSN), to estimate traffic flows under dynamic user equilibrium. In particular, HSTGSN exploits dependency between origin and destination nodes, even when it is long-range, and learns implicit vehicle route choices under different origin-destination demands. This model is based on a heterogeneous graph which consists of both the road links and OD links, which connects origins and destinations, and a spatio-temporal graph encoder-decoder that captures the spatio-temporal relationship between OD demands and flow distribution. Besides, we will show how the graph encoder-decoder is able to recover the incomplete information in the OD demand. Using extensive experimental studies on real-world networks with complete/incomplete OD demands, we demonstrate that our method can not only capture the implicit spatio-temporal relationship between link traffic flows and OD demands but also achieve accurate prediction performance and generalization capability.

109 - Does new technology complement or replace human agents? Empirical evidence from an online real estate platform in China

Yidi WANG, Communication University of China, Beijing, China, People's Republic of, Lin Wang, Zan Yang, Chong Wang

The integration of Virtual Reality (VR) technology into online housing platforms has revolutionized the real estate industry, offering innovative solutions to address information asymmetry and reduce reliance on traditional agents. Although technological advance promises to spur economic growth, there is widespread concern that it could displace workers and disrupt employment. Collaborating with a leading Chinese real estate brokerage, we analyze empirical data to investigate whether VR technology complements or replaces human agents. Our findings reveal that while VR empowers agents in temporal and spatial dimensions, there is a decline in commission earnings. Interestingly, agents with lower customer satisfaction and less experience benefit from the adoption of VR technology, leading to increased work efficiency. This highlights the changing landscape of skill requirements, where standardized hard skills such as education and certification diminish in importance, while soft skills such as customer satisfaction, initiative, and years of experience emerge as critical capabilities for agents to withstand VR substitution.

This study contributes to the literature by shedding light on the nuanced dynamics of human-computer interaction in the real estate sector, elucidating how technology both complements and threatens human roles. Additionally, it expands the discourse on VR technology's impact on real estate agents, addressing gaps in existing research on the consequences of VR implementation. Furthermore, by emphasizing the importance of soft skills in navigating the digital era, this study underscores the need for education and professional training reforms to equip human professionals with the requisite skills for future success in technology-enhanced environments.

Tuesday, October 22, 2:15 PM - 3:30 PM

APS Poster

Flex C

APS Poster Session

Poster Session

Applied Probability Society

Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Sequential Decision-Making under Preference Feedback

Jung-hun Kim, Seoul National University, Gwanak-gu, Korea, Republic of

coming soon

2 - Re-solving Policies on Finite-Horizon Restless Bandits

Chen Yan, Universitiy of Michigan, Ann Arbor, MI, United States

TBD

3 - Sequential Decision Making Algorithms with Function Approximation

Taehyun Hwang, Seoul National University, Seoul, Korea, Republic of

TBD

4 - Structural Heterogeneity and Performance in Stochastic Networks: From Distributed Inference to Epidemics and Beyond

Mansi Sood, MIT, Cambridge, MA, United States

TBD

5 - Learning, Incentives, and Adaptivity in Online Markets

Shiliang Zuo, UIUC, Savoy, IL, United States

TBD

6 - Dynamic Asset Pricing and Allocation with Expert Forecasts

Anas Abdelhakmi, IORA, National University of Singapore, SIngapore, Singapore

TBD

7 - Optimizer's Information Criterion: Dissecting and Correcting Bias in Data-Driven Optimization

Tianyu Wang, Columbia University, New York, NY, United States

tbd

8 - TBD

Anirudh Sridhar, MIT, Cambridge, MA, United States

TBD

9 - Data-driven Population Tracking in Service Systems

Morgan Wood, UNC Chapel Hill, Whitakers, NC, United States

tbd

10 - Asymptotically Optimal Adaptive AB tests for Average Treatment Effect

Vikas Deep, Kellogg School of Management, Northwestern University, Evanston, IL, United States

tbd

11 - TBD

Hanzhao Wang, Imeprial College London, London, United Kingdom

TBD

12 - Limit Theorems for Stochastic Gradient Descent with Infinite variance

Wenhao Yang, Stanford University, Stanford, CA, United States

tbd

13 - Policy Evaluation in Dynamic Experiments

Yuchen Hu, Stanford University, Stanford, CA, United States

tbd

14 - TBD

Suqi Liu, Harvard University, Boston, MA, United States

TBD

15 - Asymptotic Product-form Steady-state for Multiclass Queueing Networks in Multi-scale Heavy Traffic

Dongyan Huo, Cornell University, Ithaca, NY, United States

tbd

16 - TBD

Reese Pathak, UC Berkeley, Berkeley, CA, United States

tbd

17 - Random Graph matching with multiple graphs

Taha Ameen, University of Illinois Urbana-Champaign, Urbana, IL, United States

TBD

18 - Lyapunov methods for restless bandits

Yige Hong, Carnegie Mellon University, Pittsburgh, PA, United States

TBD

19 - Binary Perceptron: From Phase Transitions to Efficient Algorithms

Shuangping Li, Stanford University, Stanford, CA, United States

tbd

20 - Exploiting Observation Bias to Improve Matrix Completion

Yassir Jedra, MIT, Cambridge, MA, United States

tbd

21 - Dynamic Assignment of Jobs to Workers with Learning Curves

Parshan Pakiman, Chicago Booth, Chicago, IL, United States

tbd

22 - Efficient Gradient Estimation for Overparameterized Stochastic Differential Equations

Shengbo Wang, Stanford University, Stanford, CA, United States

tbd

23 - The BAR-approach for Continuous-time Multi-scale Generalized Jackson Networks and Parallel-server Systems

Yaosheng Xu, University of Chicago, Chicago, IL, United States

tbd

24 - Best of Both Worlds Guarantees for Smoother Online Quadratic Optimization

Neelkamal Bhuyan, Georgia Institute of Technology, Atlanta, GA, United States

tbd

25 - Performance of NPG in Countable State-Space Average-Cost RL

Yashaswini Murthy, University of Illinois Urbana-Champaign, Urbana, IL, United States

tbd

26 - Nearly Minimax Optimal Regret for Multinomial Logistic Bandit

Joongkyu Lee, Seoul National University(SNU), Gwanak-gu, Korea, Republic of

tbd

27 - Online Learning for Queueing Systems

Guiyu Hong, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of

tbd

28 - Principled Data Decisions in AI Ecosystems

Judy Hanwen Shen, Stanford University, Palo Alto, CA, United States

tbd

29 - tbd

Swati Padmanabhan, MIT, Cambridge, MA, United States

tbd

30 - AI and Data Analytics for Large Scale Decision-Making in Complex Dynamic Environments

Marios Papachristou, Cornell University, Ithaca, NY, United States

tbd

31 - TBD

Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

tbd

32 - Machine Learning for Operations: Bridging AI, Incentives, and Social Impact

Anand Kalvit, Stanford University, Stanford, CA, United States

tbd

33 - Learning-Based Pricing and Matching for Two-Sided Queues

Zixian Yang, University of Michigan, Ann Arbor, MI, United States

tbd

34 - Optimal Congestion Signaling to Customers with Heterogeneous Patience

Prakirt Jhunjhunwala, Columbia Business School, New York, NY, United States

tbd

TD01

Summit - 320

Service Operations and Supply Chain Management

Invited Session

Service Science

Chair: Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States

1 - Sourcing Strategies under Correlated Random Yields and Competition

Arda Yenipazarli, Georgia Southern University, Statesboro, GA, United States

This paper examines the strategic sourcing and diversification decisions of firms that procure from multiple unreliable suppliers and engage in quantity-based Cournot competition in the end market. Specifically, we analyze a two-echelon supply chain where two competing firms order a critical component from two external suppliers and use it to produce and sell substitutable products in the end market. The suppliers' production processes are unreliable, with the quantity delivered relative to that ordered modeled by correlated proportional random yields following a general joint distribution. Additionally, the two suppliers differ in terms of their per-unit procurement costs and yield distributions. We explore how these asymmetries and the correlation between suppliers' yields affect the firms' equilibrium sourcing strategies. Our findings provide insights into the conditions under which firms opt for sole-sourcing versus dual-sourcing and the resulting implications for competition and risk management in the supply chain.

2 - Assessing the Impact of Brand-Level ESG Violations on Sales

M. Serkan Akturk, Clemson University, Clemson, SC, United States, yao chen, Rakesh Mallipeddi, Arvind Mahajan

Employing retail transaction data and firm-level environmental, social, governance (ESG) information, our research investigates the impact of firms' ESG violations on their operational performance. We show that ESG violations lead to decreased sales for brands. Furthermore, store location and customer demographics moderate the relationship between ESG violations and brand sales.

3 - Learning-Based Optimization of Telemedicine Appointments

Ali Dogru, University of Southern Mississippi, Hattiesburg, MS, United States, Emirhan Uzunlar, Ibrahim Capar, Ozgur Araz

As more medical practices transition to a hybrid approach, simultaneously offering virtual and in-office consultations, scheduling these appointments remains a challenge due to complex computational requirements. We propose a learning-based stochastic optimization model to improve appointment capacity planning in such clinics. Using MEPS data, we train a machine learning model that predicts whether a patient will likely prefer a virtual visit or an in-office consultation. These predictions are then fed into the stochastic optimization model that finds an appointment scheduling template that minimizes the weighted average of patient indirect and direct waiting time, physician idle time, and overtime. The model considers service time variability, patient no-shows and unpunctuality, and same-day appointment demand. We show the utility of the proposed model using operational data from a psychiatry clinic.

4 - Quality Differentiation vs. Resource Utilization: Does Product Size Matter?

Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States, Dongling Huang, Isil Koyuncu

Using a large-scale airline dataset, we investigate how customer-perceived quality and resource consumption differences of products may impact a firm's product line decisions, and how they relate to its success. We hypothesize that (1) indirect effects of quality differentiation exist through resource allocation decisions and (2) increasing resource consumptions have negative implications.

TD02

Summit - 321

Advanced Data Analytics on Reliability and Maintenance

Invited Session

Quality, Statistics and Reliability

Chair: Vipul Bansal, University of Wisconsin Madison, Madison, WI, United States

Co-Chair: Akash Deep, Oklahoma State University, Stillwater, United States

1 - Federated Multi-Way Pca for Heterogenous Data

Madi Arabi, North Carolina State University, Raleigh, NC, United States, Xiaolei Fang

Tensor data, prevalent in multidimensional datasets, poses challenges for dimensionality reduction. Multiway Principal Component Analysis (MPCA) is a widely established method to address this by seeking projection matrices for dimensionality reduction across all modes. While traditional MPCA assumes tensors of identical dimensions, this paper extends MPCA to heterogeneous data from multiple users. A federated heterogeneous MPCA (FHPCA) framework is proposed, enabling users to maintain data privacy. FHPCA employs a collaborative Incremental Singular Value Decomposition (ISVD) approach in updating shared matrices while protecting privacy. The proposed method achieves dimensionality reduction and data privacy in heterogeneous tensor datasets, guaranteeing the same performance as traditional MPCA.

2 - Pomdp-Based Optimal Maintenance Planning Using Multiple Sensor Signals

Ahmad Salehiyan, Oklahoma State University, Stillwater, OK, United States, Akash Deep, Jaesung Lee, Minhee Kim

Effective maintenance planning is critical for optimizing system reliability. In the era of digital technologies, modern systems are equipped with multiple sensors that capture and relay information about the health of the system. However, often maintenance planning methods rely on singular data sources, such as isolated sensor readings or historical maintenance records, which may fail to provide a comprehensive understanding of equipment health and performance. This research fills this gap by proposing a framework integrating multiple sensor signals for optimal maintenance planning under partial observability. Specifically, we model the underlying degradation process as a Markov chain and relate the observable multivariate signals through a multivariate normal distribution. This modeling approach enables fusing diverse signals to estimate degradation state probabilities. We formulate the maintenance planning problem as a partially observable Markov decision process to address uncertainty from the multiple heterogeneous signals. Under certain assumptions, we prove the optimal policy possesses a control-limit structure on belief space. Extensive numerical studies and a real-world case study are used to demonstrate the approach and existence of the control limits based on cost parameters.

3 - A Scalable Approach for Condition Based Maintenance of Large K-Out-of-N Systems

Han Hu, University of Wisconsin-Madison, Madison, WI, United States, Vipul Bansal, Shiyu Zhou, Yong Chen

Condition-based maintenance(CBM) has attracted widespread attention as it can helps to reduce maintenance cost and enhance reliability of systems. For large K-out-of-N system, a common problem in CBM is that finding optimal maintenance solution is usually too computationally demanding due to the exponential explosion of state space and action space, when there is dependence between components in the system. In this work, we propose a scalable approach to find approximated solution to CBM of large-scale K-out-of-N system with economic dependence. The formulation of component-wise Markov Decision Process(CW-MDP) and adjusted component-wise Markov Decision Process(ACW-MDP) are based on extended single-component actions space and distributed single-component reward function so that component-level solutions are obtained and help to make system-level decision without using huge amount of computational resources. Convergence properties are analysed in this paper to show theoretical optimality gap. Numerical study and case study are conducted to demonstrate the effectiveness and efficiency of our approaches.

4 - Joint Model for Multi-Type Failure Event Prediction from Multi-Sensor TIME Series and Survival Data

Sina Aghaee Dabaghan Fard, Texas A&M University, College Station, TX, United States, Jaesung Lee

Modern industrial systems often have multiple components that are subject to multiple types of failure, and their conditions are monitored by multiple sensors, generating multiple timeseries signals. Additionally, time-to-failure data are commonly available. It is crucial to predict the remaining useful life (RUL) of multiple components by leveraging both multi-sensor timeseries and multi-type failure event data. In most existing models, failure mode identification and RUL prediction are performed independently, ignoring the inherent relationship between these two tasks. Some models integrate multiple failure modes classification and event prediction by using black-box machine learning approaches, which lack statistical rigor and are unable to characterize the inherent uncertainty in the system and data. In this paper, we introduce a unified approach to jointly model the multi-sensor timeseries data and RUL prediction with respect to multiple failures. We integrate a Cox proportional hazards model with a Multi-output Convolutional Gaussian Process (MCGP) in a hierarchical Bayesian framework. We employ a Dirichlet prior to account for multiple failure modes to precisely capture distinct degradation paths specific to each failure mode and accurately predict the failures within Cox proportional hazard model. Variational Bayes is used for inference, where we derive an Evidence Lower Bound (ELBO). Due to the unified framework that characterizes the associations between multiple sensor signals and multi-type failure events and their uncertainties, the proposed method outperforms existing methods.

5 - A New View of Neural Network-based Health Index: Connecting Visualization, Prediction, and Operational Decision Making

Zihan Li, University of Florida, gainesville, FL, United States, Minhee Kim

Multisensor signals are now easily accessible for remaining useful life (RUL) prediction and health state monitoring due to the quick advancement of sensor technologies. In order to optimize these signals for improved health condition assessment and RUL prediction, a variety of data fusion approaches are frequently used to generate health indices. However, most of the existing methods hardly consider the interactions between Prognostics and condition-based maintenance (CBM), which may not be effective for practitioners' asset management. To address the issue, we provide a new view of the Health Index that provides Connection Visualization between Prognostics and CBM, Prediction, and Operational Decision-Making Suggestions. In particular, a neural networks-based structure is used, and a novel loss function is developed by taking into account the variability of the generated health index at the failure time as well as its monotonicity. Once the prognostic model is trained, then we search for the best model weights that minimize the maintenance costs along “low loss curves” connecting various trained prognostic model weights. A case study on the degradation of aircraft gas turbine engines is also constructed which shows an improvement in maintenance costs of the proposed method.

TD03

Summit - 322

Advanced Machine Learning for Complex Physical Systems

Invited Session

Quality, Statistics and Reliability

Chair: Yinan Wang, Rensselaer Polytechnic Institute, Troy, United States

Co-Chair: Mucun Sun, Idaho National Laboratory, Idaho Falls, ID, United States

1 - Synth4Seg - Learning Defect Data Synthesis for Defect Segmentation using Bi-level Optimization

Shancong Mou, Georgia Institute of Technology, Atlanta, GA, United States

Defect segmentation is crucial for quality control in advanced manufacturing, yet data scarcity poses challenges for state-of-the-art supervised deep learning. Synthetic defect data generation is a popular approach for mitigating data challenges. However, many current methods simply generate defects following a fixed set of rules, which may not directly relate to downstream task performance. This can lead to suboptimal performance and may even hinder the downstream task. To solve this problem, we leverage a novel bi-level optimization-based synthetic defect data generation framework. We use an online synthetic defect generation module grounded in the commonly-used Cut\&Paste framework, and adopt an efficient gradient-based optimization algorithm to solve the bi-level optimization problem. We achieve simultaneous training of the defect segmentation network, and learn various parameters of the data synthesis module by maximizing the validation performance of the trained defect segmentation network. Our experimental results on benchmark datasets under limited data settings show that the proposed bi-level optimization method can be used for learning the most effective locations for pasting synthetic defects thereby improving the segmentation performance by up to 18.3% when compared to pasting defects at random locations. We also demonstrate up to 2.6% performance gain by learning the importance weights for different augmentation-specific defect data sources when compared to giving equal importance to all the data sources.

2 - Outage Management in Power Distribution Networks Using Reinforcement Learning over Graphs

Jie Zhang, The University of Texas at Dallas, Richardson, TX, United States, Roshni Jacob, Steve Paul, Souma Chowdhury, Yulia Gel

Self-healing smart grids are characterized by fast-acting, intelligent control mechanisms that minimize power disruptions during outages. The corrective actions adopted during outages in power distribution networks include reconfiguration through switching control and emergency load shedding. The conventional decision-making models for outage mitigation are, however, not suitable for smart grids due to their slow response and computational inefficiency. Here, we present a new graph reinforcement learning (GRL) model for outage management in the distribution network to enhance its resilience. The distinctive characteristic of our approach is that it explicitly accounts for the underlying network topology and its variations with switching control, while also capturing the complex interdependencies between state variables (along nodes and edges) by modeling the task as a graph learning problem. Our model learns the optimal control policy for power restoration using a Capsule-based graph neural network. We validate our model on two test networks, namely the 13, 34 and 123-bus modified IEEE networks where it is shown to achieve near-optimal, real-time performance. The resilience improvement of our model in terms of loss of energy is 607.45 kWs and 596.52 kWs for 13 and 34 buses, respectively. Our model also demonstrates generalizability across a broad range of outage scenarios.

3 - Neural Operator for Accelerating Coronal Magnetic Field Computations

Bo Shen, New Jersey Institute of Technology, Newark, NJ, United States, Yutao Du, Qin Li, Haimin Wang

The application of the Tensorized Fourier Neural Operator (TFNO) to significantly enhance the computational efficiency of coronal magnetic field calculations within the Bifrost Magnetohydrodynamics (MHD) model is introduced in this study. Leveraging simulated data from the European Sunrise Science Data Center, the TFNO— an extension of the Fourier Neural Operator (FNO) that incorporates tensor decomposition for improved handling of high-dimensional data—is employed to solve time-varying partial differential equations (PDEs) over a 3D domain. The performance of the TFNO is compared with traditional machine learning methods, including Vision Transformer and CNN-RNN (encoder-decoder) architectures, to demonstrate its accuracy, computational efficiency, and scalability. A physics-analysis of the TFNO predictions is also performed to demonstrate the reliability of the method. This advancement not only accelerates the simulation of solar atmospheric phenomena but also provides more reliable prediction capabilities, thus greatly contributing to the understanding of space weather dynamics and its impact on Earth.

4 - DeepMIDE: A Multivariate Spatio-Temporal Deep Learning Model for Ultra-Scale Offshore Wind Energy Forecasting

Feng Ye, Department of Industrial & Systems Engineering, Rutgers University, Piscataway, NJ, United States, Xinxi Zhang, Michael Stein, Ahmed Aziz Ezzat

The hyper-increase in scale and capacities of modern offshore wind turbines—with altitudes approaching many of the world's tallest landmarks–motivates a departure from univariate wind energy forecasting methods that are only focused on a single representative height. In response, we propose DeepMIDE—a multivariate statistical deep learning method which—unlike existing time-series and space-time methods—jointly models the spatial, temporal, and vertical variations and dependencies in the offshore wind resource and corresponding power output. DeepMIDE is a multi-input, multi-output integro-difference equation model with a nonstationary multivariate kernel, for which key physically meaningful parameters are learnt via an advanced deep transformer architecture that extracts the governing dynamics of wind propagation and formation from streams of weather maps. Tested on real-world data from the offshore wind energy areas in the U.S. East Coast—where Giga-watt scale offshore wind farms are in-development— DeepMIDE is shown to produce forecasts that are of higher accuracy and enhanced interpretability than prevalent statistical and deep learning-based forecasting methods.

TD04

Summit - 323

Survival Analysis and Optimization Techniques

Contributed Session

Chair: Sundaravalli Narayanaswami, Indian Institute of Management, Ahmedabad, India

1 - An Overview of Survival Analysis and its Use Case in Fraud Detection

Kiruthika Sankaran, Expedia, Chicago, IL, United States

Survival analysis is a crucial statistical methodology for analyzing time-to-event data, offering insights into the probability of an event occurring over time while accounting for the complexities of real-world data. This presentation will provide a comprehensive overview of survival analysis, covering its key concepts, applications, and methodologies. We will delve into the foundational elements such as survival functions, which estimate the probability of survival beyond a certain time point, and hazard functions, which describe the instantaneous risk of the event occurring. Both parametric and non-parametric survival analysis techniques will be discussed, highlighting methods like the Kaplan-Meier estimator and Cox proportional hazards model, alongside evaluation metrics such as the log-rank test and concordance index.

Furthermore, we will explore a novel application of survival analysis in the domain of fraud detection, specifically within the context of onboarding hotels onto an online platform. The process of onboarding new hotels can be susceptible to fraudulent activities, which pose significant risks to the platform's integrity and financial health. By applying survival analysis, we aim to identify patterns and risk factors associated with fraudulent behavior. This involves segmenting the hotel data into subgroups and assessing their risk profiles over time. The analysis provides valuable insights into the timing and likelihood of fraudulent events, enabling the development of proactive strategies to mitigate fraud. Through case studies and real-world examples, we will demonstrate how survival analysis can be leveraged to enhance fraud detection mechanisms, ultimately contributing to more secure and efficient operational practices.

2 - Reducing the ChvÁTal Rank Through Binarization

Vrishabh Patil, Carnegie Mellon University, Pittsburgh, PA, United States, Gerard Cornuejols

In a classical paper, Chvátal introduced a rounding procedure for strengthening the polyhedral relaxation P of an integer program; applied recursively, the number of iterations needed to obtain the convex hull of the integer solutions in P is known as the Chvátal rank. Chvátal showed that this rank can be exponential in the input size L needed to describe P. We give a compact extended formulation of P, described by introducing binary variables, whose rank is polynomial in L.

3 - A Spatial Price Network Equilibrium Paradox

Ismael Mohammad Pour, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Ladimer Nagurney

Spatial price equilibrium models have had wide application in commodity trade in sectors such as agriculture and mineral and energy markets. Many of such models are network-based and their relevance is now heightened due to global issues such as increasing wars and crises as well as the negative impacts of climate change. Spatial price equilibrium problems are network equilibrium problems, which include traffic network equilibrium problems. We investigate whether an analogue to the Braess Paradox can occur in spatial price network equilibrium problems, which, unlike the classical Braess Paradox problem, have elastic demands, since consumers are sensitive to price. We construct a series of examples in which a spatial price network equilibrium paradox occurs in that, after the addition of a new transportation link, the demand price for the commodity rises; the supply price for the commodity decreases, and all transportation routes are now more costly. In the case of agricultural commodities such a paradox has negative implications for food security, since consumers receive a lower volume of the commodity but at a higher price. Farmers, on the other hand, also, suffer, since the supply price of the commodity decreases.

4 - Impact of a Single TIME Zone on the Indian Economy

Sundaravalli Narayanaswami, Indian Institute of Management, Ahmedabad, India

Due to the Earth’s rotation, different locations on the globe experience the day at different times. Prior to time standardisation, it was difficult to coordinate activities across vast distances. To address this problem, the globe was logically divided into longitudinal regions that shared the same time, known as a time zone. Each time zone is either ahead or behind Coordinated Standard Time (UTC), which is the primary standard adopted by the world to synchronise and regulate its time. The longitudinal extent to which one time zone is ahead of another gives us a measure for how much time difference exists between the two regions. A difference of 15 degrees longitude approximates to an hour’s difference in real time.

Given this difference in time across longitudes, many countries around the world have adopted multiple time zones within their borders to better accommodate the differences in time between its own regions. For example, Russia has 11 time zones, Australia has 3, and France has 12. There are also countries that, despite spanning thousands of kilometres, have decided to retain a single time zone, say China sets its entire nation’s time to Beijing time for the sake of national unity.

India too maintains a single time zone, set at the longitude that passes through Mirzapur in Uttar Pradesh. However, on the globe, India spans across a longitudinal range of 30 degrees from east to west. In this paper, we analyze and present the economic implications of a single timezone in India.

TD05

Summit - 324

Sustainable Practices in Fashion and Commerce

Contributed Session

1 - Combating Counterfeiting in Global Supply

Seyed Mohsen Kermani, Rotterdam Managemnet School, Rotterdam , Netherlands

This research explores the effectiveness of litigation as a strategy against counterfeiting within the fashion industry. Companies face various counterfeiting challenges, including mass counterfeiting, brand dilution, and cybersquatting. They respond by filing individual or mass lawsuits against infringers. The study examines how these legal strategies impact the influx of counterfeit goods and the market dynamics, particularly focusing on collaboration among competing firms to address mass counterfeiting.

Counterfeiters adapt their strategies based on firms' legal actions, affecting their product bundling, transport modes, and packaging decisions. This research utilizes two primary datasets: detailed records of lawsuits filed by firms from 2000 to the present, and United States Customs and Border Protection data on seized counterfeit goods. These datasets provide insights into the legal battles and frontline efforts against counterfeiting.

The effectiveness of these strategies is analyzed using econometric models and theoretical frameworks such as game theory and signaling theory. The study aims to quantify the deterrent effect of litigation on counterfeiters and its overall efficacy, offering evidence-based recommendations for firms, policymakers, and enforcement agencies to enhance anti-counterfeiting measures.

2 - Attitudinally Green, Behaviorally Not: Examining the Environment Attitude Gaps through 3D Printed Clothing Choice and Diffusion of Innovation

Lakshmi Iyer, UNC Greensboro, Greensboro, NC, United States, Wei Xie

This study leverages the Diffusion of Innovation (DoI) theory to develop a Multiple Mediated Attitudes in Environmental Protection (MMAEP) research model. The model aims to understand the discrepancy between consumers' attitudinal and behavioral commitment to environmental protection by examining the adoption of 3D printing technology-enabled clothing. Based on data collected from over 221 consumers, the analysis provides valuable insight into the apparent inconsistencies in the literature regarding environmental protection attitudes and behaviors. Empirical findings reveal that consumers' environmental attitudes do not directly influence their decisions to adopt 3D-printed clothing. Instead, these attitudes interact with and exert their effects through four constructs of Compatibility, Complexity, Trialability, and Observability from the DoI theory. The study highlights the efficacy of the DoI theory in promoting environmental protection attitudes toward adopting innovations. Moreover, it introduces 'Ease of Use' and 'Imitate Others' to effectively operationalize the constructs of less straightforward 'Complexity' and conceptually ambiguous 'Observability' within the DoI framework. The proposed model enhances the understanding of how consumers' environmental protection attitudes influence consumption behaviors in the innovation adoption process. It adds evidence to explain why consumers can be ‘attitudinally green but not behaviorally green,’ contributing to the literature on Green IS for sustainability. Through its robust methodological design and empirical evidence, this study advances the theoretical and practical understanding of how environmental attitudes influence the adoption of green technologies - 3D printing.

3 - Fashion brands' second-hand business: can it have it all?

Fan SI, National university of Singapore, Singapore, Singapore, Lucy Chen

In recent years, many fashion brands have introduced trade-in programs and started to sell second-hand products through their official websites. While recycling and reusing are environmentally beneficial, it is unclear whether introducing such recycle programs is profitable. In this paper, we build a stylized model to understand the impact of recycle programs on firms’ profitability, and find that such programs can reduce a firm’s profit. Furthermore, different payment methods to second-hand product owners can affect the firm differently.

4 - An Analytic Comparison of Three Game Theoretic Models for Waste Management Supply Chains

Nima Alizadeh Basban, State Univ. of New York, Buffalo, Buffalo, NY, United States, Nallan Suresh

Waste management firms (WMFs), a key element of circular economy, form the principal means for recycling of municipal solid waste (MSW). Operationally, both competition and collaboration processes exist among WMFs. In this paper, we develop a unique waste recycling network using three main competitive models: a Cournot competition, wholesaling competition and license competition. In Cournot competition, two WMFs independently and simultaneously collect waste and establish the price of their recycled materials during the recycling process to maximize profits. In wholesaling competition, one WMF procures recyclable waste from its competitor while vying for dominance within the same market. Lastly, in license competition, a cooperative relationship is forged through a licensing arrangement between two rival WMFs. Differentiation in prices and proportions of waste sorted/composted in each model is calculated via Stackelberg game methodology and a comparison between all defined models is made to indicate conditions for maximal recycling and profitability.

TD06

Summit - 325

Stochastic Models and Networks

Contributed Session

Chair: Aditya Gopalan, UIUC, Urbana, IL, United States

1 - Scaling Limit of a Stochastic Hegselmann-Krause Model

Aditya Gopalan, University of Illinois Urbana-Champaign, Urbana, IL, United States, Partha Dey, Rasoul Etesami

We consider an infinite-dimensional stochastic Hegselmann-Krause model on the real line. In discrete time, each point of a unit-intensity simple point process independently moves halfway toward either of its left or right neighbors, chosen uniformly at random. Co-located points are merged into a single point, and the resulting simple point process is re-scaled to unit intensity. We show that when the point processes are shifted such that there is a point at the origin, there is a unique weak limit of these dynamics when the initial point process is renewal. Using a time-reversal argument, we construct a random positive measure M on R for a dual problem, and the limiting gap between the first two points is the total mass of M assigned to R. Finally, we discuss some ongoing research directions. Joint work with Partha Dey and Rasoul Etesami.

2 - Synthesis of Conditional Probability Distributions via Eigenvalue Optimization

Brent Austgen, Sandia National Laboratories, Lowell, IN, United States, John Hasenbein, Erhan Kutanoglu

In applications involving epistemic uncertainty, often only marginal probability distributions are known for populations that are presumed to have heterogeneous subpopulations. We propose a framework for synthesizing coherent conditional distributions for a set of n subpopulations under such circumstances. The core of this framework is an optimization model defined on the manifold of stochastic matrices. Geometrically, the model is akin to identifying a maximum-volume $(n-1)$-dimensional simplex that has a given barycenter and is enclosed by the $(n-1)$-dimensional standard simplex. Maximizing the volume of a simplex is equivalent to maximizing the determinant of the $n \times n$ matrix consisting of its vertices. In our model, we employ trace maximization as a linear alternative to determinant maximization. We identify the analytical form of a solution to this model, prove it is an optimum, and present necessary and sufficient conditions for it to be the unique optimum. Additionally, we show the identified optimal solution is an inverse M-matrix with its eigenvalues on the main diagonal. We apply our framework to marginal distributions of generator outages and derates estimated using data provided by ERCOT after Winter Storm Uri (2021) and demonstrate how effectively the framework is able to impose heterogeneity in the stochastic behavior of generator subpopulations.

3 - A Modified Integer L-Shaped Method for Two-Stage Stochastic Programs

Benjamin Riley, University of Minnesota, Minneapolis, MN, United States, Prodromos Daoutidis, Qi Zhang

Two-stage stochastic programs often require many scenarios, and therefore many variables and constraints, to accurately model random parameters. As a result, tailored solution algorithms that exploit the model structure are often necessary to efficiently solve problems of industrially relevant sizes.

In this work, we consider two-stage stochastic mixed-integer linear programs with relatively complete recourse and binary state variables (i.e., binary first-stage decisions that affect the second stage), which are often solved using the integer L-shaped method. A typical implementation of the algorithm for two-stage stochastic programs operates by iteratively generating optimality cuts from solutions of LP subproblems and MILP subproblems (“integer subproblems”). One drawback of the classical integer L-shaped method is that optimality cuts generated by solving integer subproblems require solving the subproblems to full optimality. This can slow the solution of problems significantly and mitigate the advantage provided by parallelization in the algorithm.

In this work, we propose a modification to the integer L-shaped method that incorporates the use of suboptimal solutions to integer subproblems while still guaranteeing convergence to an optimal solution. A multicut implementation of the algorithm is tested in two case studies. In the first case study, a two-stage stochastic program is solved to determine the optimal investment into, and relocation of, mobile production modules given uncertain time-varying demands in a supply chain. In the second case study, a two-stage stochastic program is used to determine the optimal design and scheduling of renewables-based fuels and power production networks with uncertain renewable power generation.

4 - Nonlinear cut-sharing in stochastic dual dynamic programming for log-linear autoregressive uncertainty in the right-hand side

Christian Füllner, Karlsruhe Institute of Technology, Karlsruhe, Germany, Steffen Rebennack

We consider the generation of cuts in stochastic dual dynamic programming (SDDP) for multistage stochastic linear programming problems with stagewise dependent uncertainty in the right-hand side described by a nonlinear autoregressive process. We first derive cut formulas for the general nonlinear autoregressive process. For these general cuts, the computational tractability becomes a major challenge for SDDP-type algorithms. We are, however, able to develop closed-form cut formulas for the special case of log-linear autoregressive processes. The obtained cuts are in general non-convex in the history of the stochastic process, and thus can be used to tightly approximate the occurring nonconvex value functions. As the cuts are still linear in all decision variables, they can be directly incorporated into the subproblems in SDDP without compromising their linearity. If solvers do not allow for this, our formulas can be used to adapt the intercept of a given cut to a scenario at hand in a computationally tractable way. In this sense, cuts generated for one specific scenario are shared with other scenarios. Our findings are supported by illustrative examples and by a computational study of a hydrothermal scheduling problem.

TD07

Summit - 327

Optimizing Restaurant Industry Operations

Contributed Session

Chair: Xiahua Wei, University of Washington, Bothell School of Business, Bothell, WA, United States

1 - Digital Consumption and Self Control

Zheng Wu, Purdue University, West Lafayette, IN, United States, Jiong Sun

The harmful effects of digital over-consumption become increasingly severe. The goal of our research is to analytically examine how consumers exercise self-control and how firms’ pricing schemes (subscription versus pay-per-use) and product design strategies affect their profit, consumer surplus, and consumers’ long-term health. Our findings indicate that firms’ pricing schemes and their impact on consumers’ long-term well-being are dependent on consumer characteristics and the nature of product category being consumed.

2 - Optimizing Operations in the Restaurant Industry: Evaluating the Effects of Transitioning from Tipping to Profit-Sharing

Yi Liu, University of South Dakota, Vermillion, SD, United States, Punya Chatterjee

This paper examines the impact of replacing tipping with a profit-sharing model for servers in a restaurant using an M/M/1 model. The results indicate several advantages, including improved service capacity, increased profit for the business owner, and higher income for servers. Notably, the study reveals that under certain conditions, menu prices without tipping can be lower than in the traditional tipping model, resulting in a greater customer surplus. These findings offer valuable insights into alternative pricing strategies in the restaurant industry, empowering business owners to optimize their operations for better financial performance and customer satisfaction

3 - From Scrolling to Dining: Effects of Instagram Posts on Bib Gourmand Restaurant Visitation through the Lens of ELM

Yun-An Yeh, National Cheng Kung University, Tainan, Taiwan, You-Hung Lin, Shih-Shuo Yeh, Hsin-Hsin Chang

Abstract

In the era of digital communication, social media such as Instagram have emerged as powerful platforms for restaurants to showcase their offerings and attract potential diners. Thus, understanding how users process dining information on these platforms has become imperative. This study employs the Elaboration Likelihood Model (ELM) as a theoretical framework to investigate the influence of Instagram posts on viewers' intentions to visit restaurants, with the target of Bib Gourmand restaurants. The research proposes two central route variables: Information Congruency and Information Comprehensiveness, and two peripheral route variables: Photo Presentation and Brand Identification. A total of eight scenarios are designed to represent various combinations of these variables in Instagram posts. Participants are randomly assigned into one of the eight Instagram posts of a Bib Gourmand restaurant, and are asked to complete a questionnaire assessing perceived informativeness, persuasiveness, attractiveness, and perceived restaurant image. The present study comprises seven hypotheses, all of which will be analyzed using PLS-SEM, elucidating the corresponding research findings. In light of the power of Instagram in user-generated content (UGC), the current study serves to provide a better understanding of how different characteristics of Instagram posts shape users' attitudes toward restaurants in Web 2.0 social commerce.

Keywords: Elaboration Likelihood Model; Instagram Usage; User-Generated-Content; Bib Gourmand restaurants

4 - Exploring the Relationship between Social Media Engagement and Stock Prices: Insights from Reddit

William Uyeta, University of Washington Bothell, Bothell, WA, United States, Xiahua Wei

As digital platforms increasingly influence financial markets, the role of social media engagement in shaping stock performance has gained growing attention. This study investigates the link between social media activity and stock prices, focusing on Reddit, a prominent platform for individual investors to share opinions. Analyzing daily data on the most frequently discussed stocks on Reddit, we examine how discussion volume and sentiment of investors drive stock price movements. Our findings reveal that both discussion volume and positive sentiment are significantly associated with rising stock prices. These results provide insights for investors and financial analysts, highlighting the value of monitoring social media trends to inform investment strategies.

TD08

Summit - 328

On Security and Rescue Operations

Contributed Session

Chair: Norio Hibiki, Keio University, 3-14-1 Hiyoshi, Kohoku-Ku, Yokohama, 2238522, Japan

1 - Spatio-Temporal Analysis for Fleet Tracking and Prediction Using Indicator Kriging

Esther Jose, University at Buffalo, Buffalo, NY, United States, Moises Sudit, Rajan Batta

In maritime operations, the tracking and prediction of fleet movements are paramount for various applications ranging from navigation safety to security measures. Assuming that vessels in a fleet are spatially and temporally autocorrelated, we fit different spatio-temporal covariance models to observed fleet movements. In our experiments, the separable covariance model and the simple sum metric covariance model are found to be the best fit with respect to the mean squared error. Subsequently, spatio-temporal indicator kriging is utilized to generate probability heatmaps that estimate the likelihood of finding a vessel from the fleet in a certain location based only on the current positions of other vessels in the fleet and/or the vessels' previously observed positions. This methodology can enhance fleet-tracking in scenarios where tracking data is sparse or lost, facilitating increased situational awareness. An MIP is used to determine which ships from different fleets to locate at specific times, considering resource constraints and budget limitations. The objective is to maximize the accuracy of the generated heatmaps for each fleet.

2 - Synergistic Path Planning Considering Shipborne Uavs for Maritime Emergency Search and Rescue

Min Wang, Harbin Engineering University, Harbin, China, People's Republic of, Yuzhen Hu

Traditionally, maritime emergency search and rescue (MESAR) missions are carried out by ships, which is not cost-effective and efficiency. In recent years, Unmanned Aerial Vehicles (UAVs) equipped with search sensors or relief supplies are becoming a research focus. However, short endurance of UAVs limits their ability to conduct large-scale missions. Using a ship as a mobile supply base of UAVs can make MESAR cheaper, faster and broaden its coverage. Therefore, this paper investigates the synergistic path planning problem of shipborne UAVs for MESAR. We introduce an exact approach based on branch and bound to obtain the best sequence and clustering of the SAR locations. Each cluster will be visited by a ship and a fixed UAV group, whose launch and retrieving locations will be evaluated by a second-order cone programming. We also propose several heuristics to improve solving efficiency based on the programming. Moreover, our approaches are flexible to accommodate various constraints and objective functions of the second-order cone programming. Computational results demonstrate that our proposed solution approaches can find a satisfactory synergistic path of shipborne UAV within a feasible time.

Additionally, shipborne UAV mode significantly outperforms using ships or UAVs separately to search for people injured or transport relief supplies. Deep insights are also obtained by studying the influence of some parameters on the final solution.

3 - FURIOUS: Fully Unified Risk-assessment with Interactive Operational User System for vessels

Yooyeun Kim, Seoul National University, Seoul, Korea, Republic of, Jeehong Kim, Hyunwoo Park

Over recent years, marine collision risk assessment has advanced, enhancing maritime safety. However, existing studies often describe ship domains and collision risk assessments in a static manner, lacking interactivity. Notably, interactive visualization of collision risk in multi-ship encounter scenarios has not been sufficiently developed. This gap prompted the development of "FURIOUS: Fully Unified Risk-assessment with Interactive Operational User System for vessels." This tool aids in visualizing and analyzing collision risk of multi-ship encounter situation through real-time visualization.

Our system processes Automatic Identification System (AIS) data. The backend, implemented with Python Flask, serves data to the frontend built with Next.js and TypeScript, allowing users to specify parameters such as timeframe and distance for precise navigation planning and proactive collision avoidance. Key functionalities include ship domain calculations and collision risk assessments, supported by modules for data conversion, linear interpolation of vessel lengths, and geospatial computations.

The system supports user tasks like displaying vessels, checking ship domains, and calculating collision risks. Its interactive elements and dynamic maps facilitate real-time decision-making, ensuring navigation safety. Users can interactively adjust parameters such as ship type, time window of focus, and ship IDs, enabling tailored analyses and proactive collision avoidance. Additionally, the system contributes to novice users' understanding of complicated maritime dynamics, fostering innovation in maritime operations.

Our paper details the design, implementation, and evaluation of this tool, emphasizing its potential to transform maritime decision-making by improving situational awareness, promoting navigational safety, and enhancing operational efficiency for a safer and more sustainable future.

4 - Generalized Neural Network Assisted Gnss for Precise Attitude Estimation with Carrier Phase Ambiguity Resolution

Raul deCelis, Rey Juan Carlos University, Madrid, Spain

Determining aircraft rotation requires expressing vectors in various reference frames, typically the body-fixed and earth inertial triad. Traditionally, a matrix derived from Global Navigation Satellite Systems (GNSS) sensors, utilizing a minimum of three of them, is utilized for this purpose, relying on carrier phase difference signals. However, numerous challenges such as multiple interference sources and noise can hinder the precise resolution of integer ambiguities linked with carrier phase difference calculations. To tackle these challenges, a new approach employing neural networks has emerged, effectively handling large datasets and enhancing reliability for resolving ambiguities in carrier phase. By leveraging disparities in pseudo-range and carrier phase data, training diverse neural network architectures can resolve ambiguities and accurately ascertain the orientation of GNSS sensor matrices. This methodology can be implemented either independently or in conjunction with additional attitude sensors like gyroscopic data, resulting in significantly improved overall accuracy.

5 - Impact of the Hybrid Lump Sum Plan on Longevity Risk and Claiming Age of Public Pension

Norio Hibiki, Keio University, Yokohama, Japan, Dai Izawa

Recently, managing longevity risk is a very important issue for households approaching retirement. While deferred benefits can hedge longevity risk, few people implement deferred benefits in practice. A new plan to encourage deferred benefits, called Lump Sum, has also been studied. Maurer et al.(2018) stated that the Lump Sum plan would increase the number of deferred beneficiaries. Although the Lump Sum plan incentivizes delayed claiming, it carries a larger longevity risk than the current delayed benefits (Status Quo). In this paper, we propose a Hybrid Lump Sum plan that achieves three goals: incentivizing delayed claim, mitigating the longevity risk and other disadvantages of the old-age pension system for active workers compared to the Lump Sum plan, and eliminating the non-claiming age. In this plan, the increase in the pension due to deferred benefits is divided into annuities and a lump-sum payment. Thus, this plan is a hybrid of the Lump Sum and the Status Quo. We build a household model to evaluate the longevity risk in the Hybrid Lump Sum plan, and show its effectiveness through simulation analysis. We conduct an original questionnaire survey to examine whether the beneficiaries have the incentive to delay claiming age under the Hybrid Lump Sum plan, and we employ a logistic regression model to analyze the features of deferred beneficiaries under each plan. Moreover, we analyze the overall longevity risk of the Hybrid Lump Sum plan and the Lump Sum plan and clarify the effectiveness by using the results obtained from the questionnaire survey.

TD09

Summit - 329

Data-Driven Mechanisms and Allocation Algorithms

Invited Session

Auctions and Market Design

Chair: Andres Cristi, EPFL, Lausanne, Switzerland

1 - Online Mechanism Design with Predictions

Cherlin Zhu, Columbia University, New York, NY, United States, Eric Balkanski, Vasilis Gkatzelis, Xizhi Tan

Aiming to overcome some of the limitations of worst-case analysis, the recently proposed framework of ``algorithms with predictions" allows algorithms to be augmented with a (possibly erroneous) prediction can be used as a guide. The goal is to obtain improved guarantees when the prediction is correct, called \emph{consistency}, while simultaneously guaranteeing worst-case bounds for arbitrarily wrong predictions, called \emph{robustness}. The majority of the work on this framework has focused on analysis of online algorithms augmented with predictions regarding the future input. A subsequent line of work has successfully adapted this framework to mechanism design, where the prediction is regarding the private information of strategic agents. In this paper, we initiate the study of online mechanism design with predictions, which combines the challenges of online algorithms with predictions and mechanism design with predictions.

We consider the well-studied problem of designing a revenue-maximizing auction to sell one item to strategic bidders arriving and departing over time, each with a private value for the item. We study the learning-augmented version where the auction designer is given a prediction regarding the maximum value over all agents. Our main result is a strategyproof mechanism whose revenue guarantees are α-consistent with respect to the highest value and $(1-\alpha^2)/4$-robust with respect to the second-highest value, for $\alpha\in[0,1]$. We show that this trade-off is optimal within a broad family of auctions, meaning that any α-consistent mechanism in that family has robustness at most $(1-\alpha^2)/4$. Finally, we extend our mechanism to obtain expected revenue proportional to the prediction quality.

2 - Stateful Posted Pricing with Vanishing Regret via Dynamic Deterministic Markov Decision Processes

Yangguang Shi, School of Computer Science and Technology, Shandong University, Qingdao, Shandong Province, China, People's Republic of, Yuval Emek, Ron Lavi, Rad Niazadeh

An online problem called dynamic resource allocation with capacity constraints (DRACC) is introduced and studied in the realm of posted price mechanisms. This problem subsumes several applications of stateful pricing, including but not limited to posted prices for online job scheduling and matching over a dynamic bipartite graph. Because existing online learning techniques do not yield vanishing regret for this problem, we develop a novel online learning framework over deterministic Markov decision processes with dynamic state transition and reward functions. Following that, we prove, based on a reduction to the well-studied problem of online learning with switching costs, that if the Markov decision process admits a chasing oracle (i.e., an oracle that simulates any given policy from any initial state with bounded loss), then the online learning problem can be solved with vanishing regret. Our results for the DRACC problem and its applications are then obtained by devising (randomized and deterministic) chasing oracles that exploit the particular structure of these problems.

3 - An Algorithm for the Assignment Game Beyond Additive Valuations

Christopher En, Columbia University, New York, NY, United States, Eric Balkanski, Yuri Faenza

The assignment game, introduced by Shapley and Shubik, is a classic model for two-sided matching markets between buyers and sellers. In the original assignment game, it is assumed that payments lead to transferable utility and that buyers have unit-demand valuations for the items being sold. Two important, and mostly independent, lines of work have studied more general settings with imperfectly transferable utility and gross substitutes valuations. Multiple efficient algorithms have been proposed for computing a competitive equilibrium, the standard solution concept in assignment games, in these two settings. Our main result is an efficient algorithm for computing competitive equilibria in a setting with imperfectly transferable utility and gross substitutes valuations. Our algorithm combines augmenting path techniques from maximum matching and algorithms for matroid intersection. We also show that, in a mild generalization of our model, computing a competitive equilibrium is NP-hard.

4 - Prophet Inequalities Require Only a Constant Number of Samples

Andres Cristi, Center for Mathematical Modeling, Universidad de Chile, Santiago, Chile, Bruno Ziliotto

In a prophet inequality problem, n independent random variables are presented to a gambler one by one. The gambler decides when to stop the sequence and obtains the most recent value as a reward. We evaluate a stopping rule by the worst-case ratio between its expected reward and the expectation of the maximum variable. Because of its connections with resource allocation and posted-price mechanisms, this problem has been intensively studied, and several variants have been introduced. While most of the literature assumes that distributions are known to the gambler, recent work has considered the question of what is achievable when the gambler has access only to a few samples per distribution. We provide a unified proof that for all three major variants of the single-selection problem, a constant number of samples (independent of n) for each distribution is good enough to approximate the optimal ratios. Prior to our work, this was known to be the case only in the i.i.d. variant.

Previous works relied on explicitly constructing sample-based algorithms that match the best possible ratio. Remarkably, the optimal ratios for the prophet-secretary and the free-order variants with full information are still unknown. Consequently, our result requires a significantly different approach than for the classic problem and the i.i.d. variant, where the optimal ratios and the algorithms that achieve them are known. We complement our result showing that our algorithms can be implemented in polynomial time.

5 - Competition between learn-and-earn pricing algorithms: A numerical study

Omer Saritac, London Business School, London, United Kingdom, Ali Aouad, Arnoud V. den Boer

Understanding how competitive dynamics between pricing algorithms affect social welfare is important for the design of online marketplaces. Recent empirical evidence and simulation-based studies suggest that advanced pricing technologies can be more likely to converge to supra-competitive prices in equilibrium.

In this work, we study the relationship between design properties of learn-and-earn algorithms and resulting market outcomes. We focus on a simple model of duopoly pricing with linear demand and cross-price substitution. We capture the differences in firms' pricing technology across three dimensions: (i) memory—volume of historical data---(ii) reactivity---frequency of price updates---and (iii) exploration---frequency of random price experiments. In synthetic numerical experiments, we find that insufficient exploration or asymmetric price technology lead to supra-competitive prices and lower social welfare. These findings support a potential new mechanism to explain market inefficiencies, caused by failure to estimate true demand under asymmetric technology.

TD10

Summit - 330

Innovative Strategies for Vaccine Distribution and Disease Management

Contributed Session

Chair: Zeynep Ertem, State University of New York- Binghamton, Vestal, NY, United States

1 - Insights into Antibiotic Resistance Pathways: A System Dynamics Approach

Avi Roy Chowdhury, Indian Institute of Technology Bombay, Mumbai, India, Jayendran Venkateswaran, Om Damani, Manoj Kumar Das

Antibiotic resistance is expected to be one of the major threats to human health in decades to come. It is evident that the emergence of antibiotic resistance is not a one-dimensional problem but still, we don't have a clear picture of its emergence. In this study, we use the system dynamics methodology to understand the emergence pathways and identify the critical pathways to explore them more. We exercised "Group Model Building" where experts from different domains gathered and discussed their concerns/beliefs. Based on their narration and works in literature, we developed a Causal Loop Diagram (CLD) model. The CLD model identifies how free access to AB and excessive AB use lead to ABR also AB use in food production causes the emergence of ABR. We built a Stock Flow Diagram (SFD) model to simulate different scenarios to further our understanding, enabling the evaluation of intervention strategies. In the SFD model, we consider the dynamics between the bacterial population and immunity and add antibiotics as an intervention method. An external influx is considered to account for the externally contaminated environment. To investigate the emergence of ABR, we further divide the bacterial population into sensitive and resistant bacterial populations and observe the effects of AB use. Also, we investigate the consequences of the duration of AB drug administration in the case of the emergence of ABR as well as in the bacterial population.

2 - An Analytical Assessment of the Effect of African-Produced Vaccines on the Gavi Vaccine Market

Ruben Proano, Rochester Institute of Technology, Rochester, NY, United States

The COVID-19 pandemic has highlighted significant disparities in vaccine access for low-income African countries. In 2021, the African Union initiated an effort to establish vaccine manufacturing capacity in Africa to meet 60% of the continent's demand by 2040. Despite Africa being the largest vaccine market by volume, it currently produces less than 1% of its demand domestically, and approximately 80% of vaccines purchased for Africa are bought with financial assistance, mainly from Gavi. Gavi and UNICEF have established an affordable vaccine market for low-income countries that takes advantage of economies of scale, pooled procurement, coordinated tenders, and innovative financial mechanisms.

This study assesses the impact of various African vaccine manufacturing capacity levels on the UNICEF/GAVI market. It analyzes the optimal decisions of profit-driven manufacturers who must decide how to meet Gavi vaccine demand with production made within and outside the African region. The study uses Monte Carlo simulation to evaluate a two-stage stochastic optimization model, where manufacturers determine production levels within and outside Africa considering pre-determined production levels and random demands.

The results indicate that achieving the African Union's 2040 goal and incentivizing manufacturers to maintain their production capacity would result in an additional cost between 2.7 and 11.9% of the annual GAVI/UNICEF vaccine procurement budget, which ratifies the benefits of having a self-sufficient vaccine market in Africa.

3 - A Data-Driven Compartmental Stochastic Programming Model for Optimal Resource Allocation during Disease Outbreaks

Sandra Eksioglu, University of Arkansas, Fayetteville, AR, United States, Hieu Bui, Ruben Proano, Haoming Shen

During disease outbreaks and other health emergencies, effective allocation of critical healthcare resources is crucial. This is particularly relevant in rural and marginalized communities, where variability in individuals' attitudes toward embracing disease prevention procedures and uncertainty in resource availability are prevalent. For example, vaccine hesitancy (VH), defined as reluctance and refusal to vaccinate, complicates the prediction of disease spread and affects the allocation of resources. In addition, the way resources are prioritized and allocated can influence these uncertainties, which pose challenges to public health planning. To address these challenges, we propose a compartmental stochastic programming (C-MSP) model that adaptively manages healthcare resources under uncertainty. Our model integrates epidemiological and logistical considerations, capturing the dynamic and interdependent nature of disease spread and resource allocation. We developed a case study using real-life data about COVID-19 vaccination uptake from January to May 2021 and data about healthcare resources in Arkansas, U.S. Using our numerical analysis, we highlight the impact of delaying the deployment of critical resources to public health; the impact of the initial stockpile and the flexibility in increasing the supply of critical resources to public health; and the price of ensuring equitable allocation of resources over time and space. Our approach aims to provide a robust framework for policymakers, especially in resource-limited rural areas, facilitating more informed and responsive decision making.

4 - Advanced machine learning techniques for epidemic disease modeling

Zeynep Ertem, State University of New York- Binghamton, Binghamton, NY, United States

The COVID-19 pandemic has highlighted the importance of effective surveillance and containment strategies in public health. In this study, we examined the impact of county-level mandates on COVID-19 cases pre-post to the availability of vaccinations. We also emphasize the need for a dynamic model of policy, as changes in population behavior, conditions, and vaccination levels can alter policy effectiveness. Dr. Ertem will be talking about several models on different machine learning techniques to address several real-life problems in public health management. Mainly, she will describe an operations research-based approach can help improve efficiency in public health decisions. She will describe a new forecasting algorithm that uses multiple data sources to timely and accurately predict epidemic diseases. She will talk about her hierarchical framework uses multi-linear regression to combine forecasts from multiple data sources and greedy optimization with forward-selection to sequentially choose the most predictive combinations of data sources.

5 - Vaccine Allocation in Human Contact Networks: A Reinforcement Learning Approach

Ehsan Ardjmand, Ohio University, Athens, OH, United States

This research presents an innovative method for enhancing vaccine distribution within human contact networks to combat infectious diseases. The proposed approach integrates reinforcement learning and graph neural networks using a novel Guided Twin Delayed Deep Deterministic (GTD3) policy gradient technique. Unique to the methodology is the prioritization of individuals for vaccination based on their network position and role, incorporating a regularization term in the actor's loss function. Through extensive numerical experiments across diverse scenarios, the superiority of GTD3 over random and centrality-based methods in networks with varying degree distributions is demonstrated. Furthermore, GTD3 exhibits robust performance across networks of different sizes and configurations, highlighting its adaptability and real-world applicability. These findings offer significant insights for public health policy and resource allocation, presenting a promising framework for managing infectious disease outbreaks in complex and dynamic environments.

TD11

Summit - 331

Data-Driven Crisis and Disaster Management

Contributed Session

Chair: FANG LIU, Durham University

1 - Privacy-Preserving Federated Learning for Human Activity Recognition

Setareh Kazemi Kheiri, State University of New York at Buffalo, Buffalo, NY, United States

Electrical line workers face significant occupational hazards, ranking among the top 10 most dangerous jobs in the United States. Ensuring their safety necessitates monitoring their activities without disrupting workflow. Wearable sensors offer a promising solution but raise privacy concerns. Addressing this gap, we propose a privacy-preserving framework using federated learning. A simulated lab experiment was conducted where participants performed various tasks of electrical line workers while wearing sensors on their dominant wrists. Acceleration data was collected, processed and used as an input to the federated weighted averaging method with adaptive learning rates. This method, compared to traditional machine learning approaches like k-nearest neighbors and support vector machines, yielded comparable accuracy (91%) while safeguarding participants' privacy. The federated learning framework allows for continuous optimization without compromising data privacy. Our findings demonstrate the feasibility of real-time activity recognition in hazardous environments. The implications extend beyond electrical line work. Our model can be applied in similar high-risk industries to monitor worker fatigue and prevent injuries. Moreover, it enables the development of ergonomic interventions tailored to real-time data, enhancing workplace safety and efficiency.

2 - Optimal Routing and Resource Allocation in Post-Disaster Scenarios through Multi-Objective Optimization and Multi-source Data Analytics

Zahra Hossein Pour, University of Arkansas, Fayetteville, AR, United States, Haitao Liao

The aftermath of hurricane events poses significant challenges for emergency management, particularly in maintaining effective transportation and resource allocation. This research proposes a novel framework that integrates multi-objective optimization with multimedia data analytics to enhance decision-making processes in real-time during disaster recovery. Our approach aims to simultaneously optimize multiple critical objectives: minimizing the time of response, maximizing the coverage of affected areas, and maximizing the utilization of available resources. Leveraging data from diverse sources, including social media, satellite imagery, and real-time traffic information, the framework provides a dynamic model to assess transportation disruptions and optimize routing strategies for emergency services effectively. The model incorporates machine learning algorithms to predict and adapt to changing conditions, ensuring optimal deployment of resources. This study will expand on previous efforts by incorporating a broader set of objectives and data sources, providing a more comprehensive and agile response mechanism. As a case study, the effectiveness of the proposed research framework can be validated using a recent hurricane event in which the objective is to significantly enhance the capabilities of emergency responders in areas impacted by the disaster.

3 - Fighting Fire with Data: Using SAS to Automate Identification of Fires in the Amazon.

Lavada Sharma, SAS Institute Inc., Raleigh, NC, United States

A fire burning in the Amazon to clear existing farmland is not nearly as devastating as a fire to clear newly deforested land, so the latter is what the Amazon Conservation Association concentrates its resources on identifying. But the ACA monitoring in search of these especially troubling blazes is currently through a manual and time-consuming review of fire alerts and satellite imagery. The SAS team was able to transform these two data sets into one and automate the generation of lists of new major fires, complete with precise coordinates of the locations that can be output into the ACA’s existing map application. In the end, the ACA said this automation took their science “to a whole other level.” SAS plans to continue assisting the ACA in this monitoring project throughout the upcoming fire season, which is expected to be even worse than usual. And this approach isn’t limited to the Amazon – it’s theoretically repeatable anywhere in the world.

4 - Optimizing JETP Implementation for Indonesia's Energy Transition

Kaarthigeswaran A Gnapathy, University of California, San Diego, La Jolla, CA, United States, Michael Davidson

The Indonesian Just Energy Transition Partnership (JETP) agreement has a stringent funding structure and does not include captive generation, which forms a significant share of greenhouse gas emissions in the country. A careful analysis is necessary to ascertain if the existing JETP agreements are helping or hindering Indonesia's energy transition. And what amendments are required between the funding partners and the government of Indonesia to ensure that JETP's energy transition goals are met.

To that end, we formulate a generation and transmission capacity expansion model that simulates JETP investments and provides insights into how the Indonesian power system would look post-JETP. This model will be a least-cost optimization model where we find the optimal generation investments to meet the projected demand in the major islands of Indonesia. The capacity built and retired for generation, transmission, and storage, along with the generation produced by the generators, will be the key decision variables for this model. We utilize publicly available data from the Indonesian government documents and the JETP secretariat and the model would introduce multiple scenarios

This study aims to explore a funding structure different from the original JETP and identify any future funding partners or collaborators that may help Indonesia achieve its JETP goals. This study also hopes to inform the International Partners Group (IPG) and Glasgow Financial Alliance for Net Zero (GFANZ) that funds JETP, the government of Indonesia, and electricity market participants in Indonesia about possible amendments to the JETP agreement and policy recommendations.

5 - Integrated Virtual Stockpile Pooling and on-Event Retrieval for Efficient Emergency Response

FANG LIU, Durham University, Durham, United Kingdom, Minxuan He, Jing-Sheng Song

To mitigate warehousing and spoilage costs in emergency supply stockpiling, governments often partner with private companies to distribute emergency supplies across various locations in the commercial supply chain for regular demand rotation. Commonly, a ``red line'' is assigned to each location, with inventory below this line reserved solely for emergencies. Previous literature has proposed a Virtual Stockpile Pooling (VSP) strategy that dynamically adjusts red lines among the locations based on real-time demand information, omitting the emergency retrieval process. This study introduces an Integrated Virtual Stockpile Pooling and On-Event Retrieval (VSPR) strategy that refines VSP by incorporating emergency retrieval costs into its design. We present a novel ranking method to solve VSPR's three-stage optimization process, leading to a collaborative base-stock system, a priority-based red-line reassignment, and a greedy algorithm for efficient supply retrieval. We show that VSPR strategically reserves supplies at ``key locations'', optimizing cost and response efficiency. A case study demonstrates that, compared with VSP, VSPR has the potential to reduce operational costs by up to 13.8% while enhancing service levels by up to 4.9%. VSPR's applicability extends to many other contexts, including resilient supply chain management, healthcare operations, energy and natural resource management, and military logistics.

TD12

Summit - 332

Optimization for Freight Transportation

Contributed Session

Chair: Amir-Behzad Samii, Vlerick Business School, Brussels, Belgium

1 - Untangling the global supply chain with artificial intelligence

Hiroki Oshima, Hitachi America, Ltd. , Santa Clara, CA, United States, Misaki Mito, Chi Heem Wong, Chakrabarti Arnab

The global supply chain faces challenges from multiple dimensions, including geopolitical tensions, environmental changes, and financial threats. Dealing with these risks requires a thorough understanding of both the logistical network and the interaction between the participants. Unfortunately, it is difficult to get visibility into the shipping process or the producer-supply network. In this presentation, we demonstrate three of our efforts to shed light on the problems faced by industry and develop solutions to address them. First, we build a temporal graph network to predict the production functions of international firms and the transactions between them. By doing so, we can better simulate impacts on the supply chain network. Second, we introduce “sensor-less” monitoring technology that can estimate internal conditions of shipping containers using meteorological data which obviates the need for cargo sensors. A visual demonstration of a working sensor-less prototype will be shown in the presentation. Third, we develop an event analysis framework to detect transportation disruptions in real-time, utilizing AIS data to estimate the location, date, and impact of the transportation disruption based on the vessel's behavior. Our technology is targeted towards tracking and prevention of cargo accumulation due to transport delays, which can lead to massive losses in accidents or disasters. These systems, working in concert, can assist companies to lower their cost of shipping, improve the quality of their services, and mitigate disruptions in the supply chain.

2 - Risk-Informed Stochastic Programming with Applications to Inland Flash Flooding of Roadway Freight Networks

Beau Groom, University of Tennessee, Knoxville, TN, United States, Mingzhou Jin

Annual freight volume is projected to grow by 50% over the next 15 years in the US, which will increase the importance of the supply chain to the economy and national security. This will exacerbate the impacts of climate change on the supply chain, a relationship that has not been well-defined for inland regions. More frequent high-intensity storms will lead to an increase in flash flooding, causing both short-term and long-term disruptions to the road network. Traditional decision-making frameworks are unable to deal with these high-impact, low-probability events, typically choosing to ignore them instead. Our proposed research relaxes the assumption of known probability distributions and uncertainty sets while incorporating the risk postures of decision makers by developing a novel risk-informed stochastic programming (RISP) model for decision making in the face of extreme events. The RISP model will be applied to the threat posed to road freight transportation by climate-change-induced flash flooding to identify strategic investment decisions to ensure supply chain performance in the midst of disruptions.

3 - A Multimodal Multicommodity Network Equilibrium Model with Service Capacity and Bottleneck Congestion for China-Europe Containerized Freight Flows

Xinyan Li, Shanghai Jiaotong University, Shanghai, China, People's Republic of

As a traditional way for transporting containerized cargoes, international liner shipping has long played a dominant role in the China-Europe freight transportation market. In recent years, China Railway Express has offered a new and competitive freight transportation option. To

provide a holistic computational tool that can predict and analyze the spatial distribution of containerized freight flows and suggest strategic transportation infrastructure and service expansion options, this paper develops and implements a multimodal multicommodity freight transportation network equilibrium model, based on a newly constructed supernetwork integrating international liner shipping lines, international freight rail services, and international and domestic highway networks in relevant countries. The combined transportation mode-route choice behavior of individual shippers is characterized by the multinomial logit model, where the values of time of different commodity categories are included in its utility functions. The congestion effects arising from bottleneck facilities along liner shipping and freight rail lines are first captured by queueing models and then approximated by polynomial delay functions, while the accommodation capacity of individual service lines is explicitly embedded into the model as hard constraints. This supernetwork equilibrium model is tackled by a subgradient algorithm in the Lagrangian relaxation framework embedding a disaggregate simplicial decomposition algorithm. Computational experiments are conducted with a multi-layer, hetero-structure set of supply and demand data. The preliminary results exhibit the computational performance of the solution procedure and reveal the variation of congestion levels of bottleneck facilities, capacity usage profiles of service lines, and competitiveness conditions between liner shipping and freight rail services.

4 - Physical Asset Criticality Index in Energy Supply Chains

Amir-Behzad Samii, Vlerick Business School, Brussels, Belgium, Raf Bellers

Disruptions can interfere with the pace of the energy transition, delaying decarbonization and other urgent actions against climate change. As a critical player in the energy supply chain, the Distribution System Operator (DSO) must fortify its operations against endogenous and exogenous supply risks, look for ways to compile knowledge and build relevant indicators to prompt an adequate response strategy. Thus, the talk focuses on deriving an asset criticality index tailored for DSO operations from a forecasting and inventory management perspective. By providing a quantitative measure of the criticality of each asset, the index allows the DSO to fine-tune internal processes and prioritize resources based on their strategic importance, ensuring that high-priority assets receive the necessary attention. Regarding inventory management, critical assets may require higher safety stock levels or closer monitoring to prevent disruptions.
The proposed asset criticality index includes several risk dimensions. We started by identifying key parameters: availability of alternative suppliers for each stock-keeping unit, the lead time, and on-time in full delivery performance that contribute to the supply risk dimension of the index. Following discussions with key experts for the DSO, we decided to what extent these parameters, and possibly others, represent risk drivers for the DSO operations and how they contribute to the index. The criticality index can also be expanded to include other vulnerabilities or risk dimensions specific to the energy supply chain, such as raw material risks, ensuring that the critical index safeguards against disruptions and actively contributes to the low-carbon transition.

TD13

Summit - 333

Strategic Innovations in Logistics and Supply Chain Management

Contributed Session

Chair: Saqib Shahzad, China, Si Chuan Sheng, Cheng Du Shi, Qing Yang Qu, Shuncheng Ave, 252号顺吉大厦10层B1-B2 邮政编码, 610017

1 - Balancing Prepositioning and Relocation for Efficient Compound Flood Relief

Muhammad Naiman Jalil, College of Business and Economics, United Arab Emirates University, Al Ain, United Arab Emirates, Syed Tariq, Muhammad Adeel Zaffar

Floods affect more people every year than any other natural disaster. Compound flooding is when multiple flooding events exacerbate the impact and duration of a flood. This paper studies relief logistics for compound flooding through the case of floods in Sindh province of Pakistan in 2010. The problem is treated as time-varying flows of relief goods from relief warehouses through periodic relief caravans to relief camps or storage close to them; and flows of affected people from shelter sites closest to them to more accessible shelter sites. Results highlight factors that reduce logistics cost and improve demand fulfillment including prepositioning relief goods near shelters, affected people relocation, frequent relief deliveries, and smaller storage modules that reflect relief demand between consecutive relief deliveries more closely. Comparative results indicate that relocation of affected people can be as cost effective as prepositioning of relief goods depending on optimal placement of relief warehouses. In case of no prepositioning, affected people relocation becomes an instrument of choice compared to relief goods transportation on longer routes during compound flooding. Temporal analysis show that affected people relocation is most effective in early stages of compound flooding, whereas goods transportation is preferred in later stages. Overall results suggest positioning of relief goods is most effective in the preparedness phase of disaster relief, relocation is most effective in the early response stage and goods transportation takes up most of the logistics activities towards late response.

2 - Boeing 787 Dreamliner Program Supply Chain Assessment

Zhiyuan Li, Southeast University, Nanjing, China, People's Republic of

This paper examines the supply chain dynamics of the Boeing 787 Dreamliner program, focusing on its global outsourcing and multi-level integration. It evaluates the effectiveness, efficiency, and feasibility of this supply chain model, highlighting the need for enhancements in product development and information transfer efficiency. Recommendations include the establishment of a lean supply chain management model, implementing supplier performance evaluation and incentive mechanisms, creating a global collaborative information platform, fostering collaboration with competitors, involving experts in supplier design and production, and prioritizing independent research and development alongside intellectual property protection.

3 - A Framework to Organize Public Transport-based Crowdshipping

Meijing Zhang, Singapore University of Technology and Design, Singapore, Singapore, Lynette Cheah

Enabled by the steady growth of e-commerce, there is a surge in demand for parcel shipments in many cities. With increasing calls for sustainable freight transport, logistics service providers are facing the need to address these challenges. Crowdshipping presents an innovative solution by involving members of the public to fulfill parcel deliveries, ideally along their pre-committed journeys. This paper aims to propose a comprehensive framework to organize public transport-based crowdshipping by integrating the use of parcel lockers. This approach allows public transport passengers to serve as crowdshippers, ensuring that parcel delivery demand aligns with crowdshipper availability and willingness to participate, while considering locker capacity constraints. The framework involves formulating and solving assignment and vehicle routing problems, incorporating matching for parcels and lockers, as well as using an iterative heuristic algorithm for optimal locker selection. The feasibility of the framework is demonstrated using real-world data, revealing significant benefits. By comparing a carrier's performance with and without engaging crowdshippers, the results indicate that delivery vehicle kilometers traveled and associated air emissions can be reduced by up to 30%. This reduction only requires redirecting less than 9% of parcels to crowdshipping. Overall, crowdshipping utilizing public transport has strong potential to be a sustainable method for fulfilling urban logistics in densely populated cities.

4 - Brand Authenticity as a Driver of Brand Loyalty in Frozen Foods sector: An Investigation of Brand Involvement and Customer Satisfaction as Mediators

SAQIB SHAHZAD, Sichuan University, Chengdu, China, People's Republic of, Adnan Sarwar, Li Shan

Purpose: The purpose of this study is to investigate the effect of brand authenticity on consumer brand loyalty in the Pakistani frozen food industry sector in the light of Stimulus Organism Response (SOR) Theory.

Design/Methodology/Approach: The quantitative approach utilized a survey questionnaire to acquire the perceptions of customers. A simple random technique was used to collect data. A total of 255 questionnaires were analyzed, and the response rate was 46.36%.

Findings: A structural and measurement model was constructed through Smart PLS-4. The findings indicated that brand authenticity has a positive effect on brand loyalty in the Pakistani frozen food industry. Results further proved that brand involvement has a full mediation effect, and customer satisfaction has a partial mediation effect.

Originality/value: This study investigates the impact of brand authenticity on brand loyalty in the frozen food industry in Pakistan. It specifically examines the mediating effects of brand involvement and customer satisfaction, which have not been extensively studied in previous research.

Keywords: brand authenticity, brand loyalty, SOR theory, brand involvement, customer satisfaction

5 - Empowering SMEs: Harnessing Collaborative Capability Marketplaces and AI for Supply Chain Transformation

Fatima Gillani, Nottingham Business School, Nottingham Trent University, Nottingham, United Kingdom, Zeng Fan, Xiao Ma, Jay Bal

Small and Medium Enterprises (SMEs) are pivotal to the global economy, yet they frequently face challenges in the rapidly changing business landscape and ongoing global crisis to scale their operations and penetrate new markets due to rigid B2B supply chain dynamics. In this realm, SMEs and their supply chains need new ways of operating that can ensure their survival and competitiveness. This study introduces a competence matching platform named the Collaborative Capability Marketplaces (CCM), which integrates advanced digital AI platforms, representing a significant paradigm shift, facilitating SMEs' connections with potential partners and engagement in larger projects. Employing a structured analytical approach, the impact of this AI-driven e-marketplace is evaluated through data from seven case studies. The findings identify two key mechanisms—competence malleability and combination—through which firms can improve their adaptability and gain a competitive edge in the marketplace. These mechanisms empower SMEs to navigate the complexities of modern markets with greater agility and foresight, allowing them to continually evolve their business models and operational tactics. The CCM acts as a transformative platform and enables SMEs to effectively showcase their competencies and access a broader range of project opportunities. The integration of Natural Language Processing aids in more accurate competency profiling and enhances the identification of suitable public and private projects; AI-based competence matching in the platform supports the development of new supply networks capable of managing high-value projects. These advancements strengthen SMEs' market positions by aligning their resources more closely with their strategic goals and market demands.

TD14

Summit - 334

Optimization and Pricing Models for Digital and Retail Markets

Contributed Session

Chair: Jisoo Park, Georgia Institute of Technology, Atlanta, GA, United States

1 - AI-Enabled Smart Retail Solutions for Improving Customer Experience Management

Jongsawas Chongwatpol, NIDA Business School, National Institute of Development Administration, Bangkok, Thailand

The challenge for physical retail stores is to offer unique value that sets them apart from the convenience of online shopping. How can AI solutions be designed and implemented effectively to manage and support the in-store customer journey and to understand the role of the customer experience at each stage of the journey? This study presents a design journey for AI implementation to understand and improve the in-store customer shopping experience by proposing a computer vision AI-based track and traceability framework with motion and emotion analysis. This study adopts the design thinking process to ensure that any challenges and failure factors in AI solutions are tackled in the proof-of-concept stage, which focuses on generating and testing the ideas. The case, presented in this study, leverages, and integrates the wealth of information made available by face recognition technologies into marketing strategies for personalizing advertisements and video content, tracking marketing and sales leads, tracking customers' motions and emotions, automating face recognition payment, and personalizing follow-up campaigns based on customers' preferences and in-store shopping habits. The contribution of this study is the proposed computer vision-AI based track and traceability system with motion and emotion analysis, which allows store managers or marketing analysts to understand customer preferences and purchasing behaviors by improving information visibility by tracking customers' movements and activities in the store.

2 - Assortment Planning Optimization under Uncertainty with Customer-driven Product Substitution

Jisoo Park, Georgia Institute of Technology, Atlanta, GA, United States, Walid Klibi, Benoit Montreuil

The evolving retail landscape, marked by increasing production costs and resource constraints, highlights the need for assortment planning decisions made with supply chain considerations. Traditionally, retail product manufacturers have assortment decisions driven by marketing, yet these decisions are intricately linked to operational aspects such as procurement, production, and inventory planning. Given the challenges posed by the unpredictable dynamics of today's retail market, we propose an integrative optimization model to address the assortment planning problem in a make-to-stock environment. We estimate customer substitution behavior with a closed-form expression based on the multinomial logit model and leverage these behaviors from the procurement and manufacturing stages to mitigate supply and demand uncertainties. Our framework addresses the challenge of aligning product offerings with supply chain capabilities and customer preferences to maximize expected profit while considering procurement and production capacity constraints. We formulate the problem as a multi-stage stochastic program with uncertain and time-varying demand, represented by multi-period look-ahead forecasts, as commonly observed in practice. Results are presented using data from an e-commerce furniture manufacturer and retailer.

3 - Co-creation and Pricing Issues in 3D Printing-as-a-Service

Jishnu Hazra, Indian Institute of Management Bangalore, Bangalore, India, Tarun Jain, Ram Gopal

The advent of 3D printing technology has facilitated new avenues for collaborative product design endeavors between manufacturers and customers. The manufacturers now leverage the 3D Printing-as-a-Service (3DaaS) model by renting out 3D printers. In this paper, we attempt to address the following research questions: What pricing model is suitable for offering 3DaaS? How do factors such as the degree of design customization and complexity impact the pricing strategy employed by the 3DaaS firm?

4 - Gig Outsourcing by Freelancers in Online Labor Markets

Sambit Tripathi, Portland State University, Portland, OR, United States, Prasanna Karhade, Amit Deokar

We explore the outsourcing strategies of freelancers in online labor markets. Literature on online labor markets considered buyers and freelancers as separate entities with specific characteristics. Studies primarily focus on how buyer attributes, freelancer attributes, and gig attributes can efficiently match freelancers and buyers. Yet, online labor markets enable buyers and freelancers to access the platform as users. Users can switch roles between buyer and freelancer in the platform. In this study, we analyze the buying and freelancing activities of the same user in an online labor market to establish the motivations of freelancers for switching to buyers. This phenomenon provides opportunities for freelancers to act as buyers and outsource specific tasks to other freelancers on the platform.

We capture users' buying and freelancing activity from a global freelancer marketplace. Next, we perform a natural language processing method to determine the similarity in the tasks completed and tasks outsourced by the same user. Our results show that freelancers conduct outsourcing strategies by hiring other freelancers to do their core and peripheral activities. This study contributes to the literature on online labor markets by introducing freelancers as buyers and explaining their outsourcing strategies. Practical implications of our research include addressing how hiring decisions are made by freelancers while outsourcing tasks to other freelancers.

TD15

Summit - 335

Analysis of Algorithms for Pricing and Product Ranking

Invited Session

Revenue Management and Pricing

Chair: Joline Uichanco, University of Michigan, Ross School of Business, Ann Arbor, MI, United States

Co-Chair: Zijin Zhang, Ross School of Business, University of Michigan, Ann Arbor, MI, 48104, United States

2 - Can Price Discrimination be Progressive?

Enfeng Xing, Tianjin University, Tianjin, China, People's Republic of, Michael Hamilton, Max Biggs, Titing Cui

Third-degree price discrimination, the practice of selling similar goods and services at different prices across various market segments, is a well-studied strategy aimed at maximizing firm revenue. The common view of price discrimination holds that it is detrimental to consumers. Price discrimination, however, can in some instances enhance consumer welfare. In this work we re-examine the impact of price discrimination on consumer welfare, paying special attention not just to aggregate gains in surplus but looking specifically at which customers in the market benefit. We highlight this by studying revenue-maximizing segmentation and price effects through the lens of proportional customer surplus which is the percentage gain in utility for a customer relative to their willingness-to-pay. Under proportional surplus, gains made by the least well-off in the market are given additional weight making it a more progressive measure of redistribution. For the case of linear demand, we show that under the revenue-maximizing segmentation and pricing, the proportional surplus increases for modest levels of price discrimination. We then generalize these results beyond linear demand, suggesting that modest levels of price discrimination can often improve relative welfare in the market even if in aggregate the total surplus decreases.

3 - Designing Pandora's Box: Personalized Rankings in Two-sided Marketplaces

Shreyas Sekar, University of Toronto - Rotman School of Management, Toronto, ON, Canada, Yiangos Papanastasiou

With the growth of e-commerce, rankings and recommendations have emerged as one of the primary levers used by two-sided marketplaces to match demand and supply. Much of the work on this topic has centered around constructing personalized rankings aimed at maximizing specific objectives (e.g., revenue, purchase probability), often assuming that the supply of sellers is fixed. However, these approaches do not account for: a) consumers' incentives to inspect products in the same order as the ranking; b) the impact rankings have on sellers' revenue and their decision to enter the platform.

In this work, we present a comprehensive model of product rankings that account for the incentives on both sides of the market. We show that traditional greedy or personalized ranking policies that myopically optimize revenue end up being sub-optimal as niche sellers are better off not joining the platform. Instead, we present randomized, incentive-compatible product ranking policies that achieve the first-best outcome by balancing the mix of mainstream and niche sellers in the top positions. Overall, our work presents a new lens to examine product ranking policies in two-sided marketplaces, and emphasizes the market-building nature of product rankings.

4 - Leveraging Offline Data for Online Decision-Making in Bayesian Multi-Armed Bandits

Wenxin Zhang, Columbia University, NEW YORK, NY, United States, Will Ma, Santiago Balseiro

This paper investigates using offline data for online sequential decision-making in Bayesian multi-armed bandit problems. We study a finite-horizon model with Bernoulli rewards, where the reward probabilities of each arm are independently drawn from unknown prior distributions. The offline dataset contains n MAB instances, each of which is a sequence of T arms pulled by some historical policies and their associated reward observations. We assume the prior is shared across the historical instances and the online instance. Empirical risk minimization is infeasible because of the lack of counterfactuals and the large size of the policy class. We introduce a dynamic programming formulation designed to derive the optimal Bayesian policy, where the expected reward of any policy is expressed as a linear combination of expectations of functions of arm reward probabilities. We then replace these theoretical expectations with unbiased estimators constructed from the data. We show the DP is solvable in polynomial time w.r.t. T, and the cumulative reward of the resulting policy converges to the optimal Bayesian policy at rate $\tilde{O}(n^{-1/2})$. Our algorithm allows a broad class of historical policies, so long as they “stay on the winner” and initially play all arms with positive probability.

TD16

Summit - 336

Innovation in Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Menglong Li, City University of Hong Kong, Kowloon, N/A, Hong Kong

Co-Chair: Venus Lo, City University of Hong Kong, Kowloon, N/A

1 - Consistent Assortment Optimization Under the Mnl with Exogenous Data

Carlos Cardonha, University of Connecticut, Storrs, CT, United States, Aritanan Gruber

This work investigates the Assortment Optimization Problem under the Multinomial Logit choice model in scenarios where the firm does not know the attractiveness factors associated with each product. In our model, the firm relies on exogenous data consisting of the assortments offered by other firms and random variables defining a set of conversion rates, which are multiplicative factors indicating how the products' attractiveness adopted in the external assortments differ from the respective values for the firm. The notion of optimality is nuanced in this model, so we focus on consistent assortments instead, i.e., we wish to make assortment decisions that do not conflict with exogenous data. We investigate different solution techniques for the problem based on assumptions about the conversion rates. We show that the expectation-based deterministic formulation of the problem can be solved efficiently. We also study a fully-separable probabilistic-constrained formulation, showing that such problems can be cast as second-order cone programs when the conversion rates are normally distributed. Our results show that the model does not require unrealistically large amounts of exogenous data to deliver accurate results, thus providing evidence that our framework can be adopted in practice.

2 - Multi-Product Dynamic Upgrades

Xiao Zhang, Saint Louis University, Saint Louis, MO, United States, Justin Goodson, Tiancheng Zhao

Upgrades in travel industry are often static and offered either at the booking time or at the check-in time. In this paper, we study dynamically-offered upgrades by a multi-product firm via notifications (e.g., emails and push notifications) between the booking and the check-in times. We investigate a general multi-level upgrade policy in which a customer may be upgraded to any better products and two single-level upgrade heuristics which are less computationally intensive. Both heuristics have clean structures respectively and are easy to implement. We adopt discrete convex analysis concepts into our analysis and identify monotonicity properties for the optimal single-level upgrade policies.

3 - Pricing Analytics with Shape-Restricted Demands

Jingren Liu, National University of Singapore, Singapore, Singapore, Hanzhang Qin, Mabel C. Chou

We consider a fundamental problem in revenue management: feature-based pricing, where a firm needs to make a pricing decision to maximize the expected revenue for a single product based on feature information. Historical prices, covariates, and sales are available for demand estimation. Our model assumes a linear relationship between price and demand while simultaneously capturing the impact of covariates through a nonparametric shape-restricted function. We develop a Three-Step Semi-Parametric Estimation algorithm to estimate the demand and foster provably near-optimal data-driven pricing decisions. From a non-asymptotic perspective, we derive finite sample regret bounds, showcasing the efficacy of our algorithm in achieving near-optimal revenue, even under potential misspecification of the demand model. The numerical results demonstrate that the decision performance of our algorithm is comparable to the Double Machine Learning method, while significantly outperforming a naive two-step iterative learning method as well.

4 - Assortment Planning in the Presence of Buy Online, Pick-up in-Store

Venus Lo, City University of Hong Kong, Kowloon Tong, Hong Kong

We consider an assortment planning problem in an on omnichannel operations where customers can purchase online and pick-up in-store (BOPS) shortly after placing an order. Retailers like to offer BOPS because customers may purchase add-on products upon when they arrive to pick up their orders and they do not need to ship products directly to customers. However, the retailer also incurs a cost for offering products under BOPS because employee manpower is needed to pick out these orders from the shelves. Hence, BOPS could be unprofitable if customers do not purchase add-on products or if the cost to prepare these orders are high. We present an approximation algorithm to show how a retailer can choose a good assortment to designate as BOPS in an online store so that he can attract customers to the physical store and maximize profitability.

TD17

Summit - 337

Online Learning in Games and Auctions

Invited Session

Revenue Management and Pricing

Chair: Christian Kroer, Columbia University, New York, NY, United States

Co-Chair: Rachitesh Kumar, Amazon, Seattle, WA

1 - Improved Rates for No-Regret Learning in General Games

Gabriele Farina, MIT, Cambridge, MA, United States

A recent line of work has established uncoupled learning dynamics such that, when employed by all players in a game, each player's regret after T repetitions grows polylogarithmically in T, an exponential improvement over the traditional guarantees within the no-regret framework. However, so far these results have only been limited to certain classes of games with structured strategy spaces---such as normal-form and extensive-form games. The question as to whether O(polylog T) regret bounds can be obtained for general convex and compact strategy sets---which occur in many fundamental models in economics and multiagent systems---while retaining efficient strategy updates is an important one. After recalling recent advances, in this talk we answer the question in the positive by establishing uncoupled learning algorithm with O(log T) per-player regret in general convex games, that is, games with concave utility functions supported on arbitrary convex and compact strategy sets. Our learning dynamics are based on an instantiation of optimistic follow-the-regularized-leader over an appropriately lifted space using a self-concordant regularizer that is, peculiarly, not a barrier for the feasible region. Further, our learning dynamics are efficiently implementable given access to a proximal oracle for the convex strategy set, leading to O(loglog T) per-iteration complexity. Even in those cases where prior results apply, our algorithm improves over the state-of-the-art regret bounds either in terms of the dependence on the number of iterations or the dimension of the strategy sets.

Based on joint work with Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Paper: https://openreview.net/pdf?id=SiSv_XDMksL

2 - Smoothed Nash Equilibria: Algorithms and Complexity

Abhishek Shetty, University of California, Berkeley, Berkeley, CA, United States, Konstantinos Daskalakis, Noah Golowich, Nika Haghtalab

A fundamental shortcoming of the concept of Nash equilibrium is its computational intractability: approximating Nash equilibria in normal-form games is PPAD-hard. In this talk, inspired by the ideas of smoothed analysis, we present a relaxed variant of Nash equilibrium called σ-smooth Nash equilibrium, for a smoothness parameter σ. In a σ-smooth Nash equilibrium, players only need to achieve utility at least as high as their best deviation to a σ-smooth strategy, which is a distribution that does not put too much mass (as parametrized by σ) on any fixed action. We show that σ-smooth Nash equilibria have superior computational properties to Nash equilibria. We will present various algorithmic results that establish that σ-smooth Nash equilibrium can be computed in polynomial time, both in terms of query and computational complexity. These results stand in contrast to the optimal algorithm for computing ϵ-approximate Nash equilibria, which cannot run in faster than quasipolynomial-time. We will also complement our upper bounds by showing that when either σ or ϵ is an inverse polynomial, finding a weak ϵ-approximate σ-smooth Nash equilibria becomes computationally intractable.

3 - Learning Dynamics in Auction Games

Martin Bichler, Technical University of Munich, Garching B. München, Germany

Equilibrium problems in Bayesian auction games can be described as systems of differential equations. Depending on the model assumptions, these equations might be such that we do not have a rigorous mathematical solution theory. The lack of analytical or numerical techniques with guaranteed convergence for the equilibrium problem has plagued the field and limited equilibrium analysis to rather simple auction models such as single-object auctions. Recent advances in equilibrium learning led to algorithms that find equilibrium under a wide variety of model assumptions. We analyze first- and second-price auctions where simple learning algorithms converge to an equilibrium. The equilibrium problem in auctions is equivalent to solving an infinite-dimensional variational inequality (VI). Monotonicity and the Minty condition are the central sufficient conditions for learning algorithms to converge to an equilibrium in such VIs. We show that neither monotonicity nor pseudo- or quasi-monotonicity holds for the respective VIs. The second-price auction's equilibrium is a Minty-type solution, but the first-price auction is not. However, the approach allows us to get ex-post guarantees for gradient-based algorithms. We show that the Bayes--Nash equilibrium is the unique solution to the VI within the class of uniformly increasing bid functions, which ensures that gradient-based algorithms attain the equilibrium in case of convergence, as also observed in numerical experiments.

4 - Robust Budget Pacing with a Single Sample

Rachitesh Kumar, Columbia University, New York, NY, United States, Santiago Balseiro, Vahab Mirrokni, Balasubramanian Sivan, Di Wang

Advertising is the economic engine of the internet. Online advertising opportunities are predominantly sold through real-time auctions: whenever a user visits the platform, an auction is run among interested advertisers, and the winner gets to display their ad to the user. Motivated by online advertising, we develop data-driven algorithms for bidding in repeated auctions under a global budget constraint. We study a non-stationary stochastic model of sequential auctions, which despite immense practical importance has received little attention, and propose a natural algorithm for it. With access to just one historical sample per auction/distribution, we show that our Dual FTRL algorithm attains (nearly) the same performance as that possible under full knowledge of the distributions, while also being robust to distribution shifts between the sampling and true distributions.

TD18

Summit - 338

Advances on Choice Modeling and Applications in RMP

Invited Session

Revenue Management and Pricing

Chair: Chenxu Ke, Shanghai University of Finance and Economics, Shanghai, MD, China, People's Republic of

Co-Chair: Ruxian Wang, Johns Hopkins Carey Business School, Kensington, MD, 20895, United States

Co-Chair: Zifeng Zhao, Mendoza College of Business, University of Notre Dame, Notre Dame, IN, United States

1 - Oligopolistic Competition in Online Marketplaces: Equilibrium Analysis and System Coordination

Xinyi Zhou, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Ruxian Wang, Guillermo Gallego, Lijian Lu

Two-sided marketplaces have grown rapidly over recent years and are ubiquitous in daily life. One important challenge is to design the marketplace to support the sustainability and efficiency of the ecosystem. This paper investigates the roles of selling formats in a two-sided marketplace with many sellers selling substitutable products on a common retailing platform. We formulate the strategic interactions between sellers and the platform in a supply function competition framework, in which each seller offers a schedule of quantity and payment that is a general function of quantities. We show that a contribution-based model (CBM), whereby the payment for each seller is based on the seller's contribution, is the only equilibrium for the supply function. We also show that the CBM results in a coordinated system achieving the first-best outcome, which is not attained under the other two widely observed schemes, reselling and agency selling. We propose a variant of CBM that yields a stable, economical, coordinated, and individually aligned outcome for all firms in the marketplace. Our findings could provide useful guidance for the design of strategic partnerships between firms in a two-sided marketplace.

2 - Room Pricing with Interdependent and Heterogeneous Price Sensitivities

Svetlana Riabova, University at Buffalo, Buffalo, NY, United States, Andrew Vakhutinsky, Natalia Kosilova, Jose Walteros

We present a data-driven approach to revenue management in the hospitality
sector. Our developments are focused on designing optimal hotel pricing policies in the presence of confounders, heterogeneity in price sensitivities, and interdependencies between the prices of different room types. The proposed approach leverages recent developments in Double/Debiased Machine Learning to develop a predictive model aimed at obtaining unbiased treatment effect estimates in the presence of confounders. We integrate the resulting predictive model into an optimization framework to determine optimal pricing policies. We tested our approach using a case study designed to emulate real scenarios that consider critical factors such as seasonality and heterogeneous customer preferences and show that the proposed model can efficiently identify the effects of the pricing policies on customer booking decisions.

3 - Contextual Dynamic Pricing: Algorithms, Optimality, and Local Differential Privacy Constraints

Zifeng Zhao, Mendoza College of Business, University of Notre Dame, Notre Dame, IN, United States

We study the contextual dynamic pricing problem where a firm sells products to T sequentially arriving consumers that behave according to an unknown demand model. The firm aims to maximize its revenue, i.e. minimize its regret over a clairvoyant that knows the model in advance. The demand model is a generalized linear model (GLM), allowing for a stochastic feature vector in $\mathbb R^d$ that encodes product and consumer information. We first show that the optimal regret is of order $\sqrt{dT}$, up to a logarithmic factor, improving upon existing upper bounds in the literature by a $\sqrt{d}$ factor. This optimal rate is materialized by two algorithms: a confidence bound-type (supCB) algorithm and an explore-then-commit (ETC) algorithm. A key insight of our theoretical result is an intrinsic connection between dynamic pricing and the contextual multi-armed bandit problem with many arms based on a careful discretization. We further study contextual dynamic pricing under the local differential privacy (LDP) constraints. In particular, we propose a stochastic gradient descent based ETC algorithm that achieves an optimal regret upper bound of order $d\sqrt{T}/\epsilon$, up to a logarithmic factor, where $\epsilon>0$ is the privacy parameter. The regret upper bounds with and without LDP constraints are accompanied by newly constructed minimax lower bounds, which further characterize the cost of privacy. Extensive numerical experiments and a real data application on online lending are conducted to illustrate the efficiency and practical value of the proposed algorithms in dynamic pricing.

4 - Product Display Optimization with Quality Uncertainty and Return

Yang Lu, University of Science and Technology of China, Hefei, China, People's Republic of, Ruxian Wang, Xiaoming Yan, Yugang Yu

Display quality refers to product information such as the images and textual descriptions of product quality presented to consumers shopping online, and can directly influence market demand during the sales phase and product returns during the after-sales phase. In this paper, we investigate the product display optimization problems on display quality, pricing, and assortment planning, in consideration of product quality uncertainty and consumers' return behavior. We employ the widely used multinomial logit (MNL) model to characterize the impact of display quality on consumer choices. In particular, we incorporate display quality into the utility function under the MNL model and characterize online consumers' purchase and return behavior that is largely affected by display quality. We then conduct joint optimization analysis on assortment planning, display quality, and pricing. Comparative statics elucidates the relationship between pricing and display quality decisions. Surprisingly, we reveal that higher product price does not necessarily lead to higher display quality unlike the classical MNL model. Then we propose an efficient method to determine the optimal offer set that involves assessing whether the conservative retention rate of a product exceeds 0. Because of its high impact on both early market demand and later consumer return behavior, display quality alters the pricing and assortment planning structure of the retailer. Moreover, we find that optimizing display quality presents an opportunity to achieve a win-win outcome for both the retailer and consumers, which could serve as a reference for market regulation.

TD19

Summit - 339

Humanitarian Logistics and Disaster Preparedness

Contributed Session

Chair: Alex Paparas, Eastern Washington University, 2603 E Keystone Ct, Spokane, 99223, United States

1 - Enhancing Refugee Resettlement: An Integrated Decision-Making Approach for Assignment

Jesica de Armas, Universitat Pompeu Fabra, Barcelona, Spain, Valérie Belanger, Marie-Ève Rancourt, Gilbert Laporte

Over the past few years, there has been a notable surge in asylum applications within the European Union (EU), underscoring the urgency for effective decision-making in the resettlement process. This study delves into the predominant outcome for asylum seekers—attaining refugee status—and the subsequent resettlement process that facilitates their transition from an asylum-providing nation to a host country. Successful integration stands as a pivotal factor in ensuring favorable outcomes for both refugees and host nations alike. Achieving this requires concerted efforts from all stakeholders, encompassing the adaptability of refugees and the preparedness of host communities. Through the National Integration Evaluation Mechanism project, integration outcomes for refugees across 15 EU member states are comprehensively assessed using a set of integration indicators. The resettlement process is further influenced by the priorities of host countries and the preferences of refugees. However, a comprehensive optimization framework that amalgamates these diverse dimensions for decision-making support is currently lacking. To address this gap, this research introduces a multiobjective decision-making system that incorporates factors such as integration potential, host countries' priorities, and refugees' preferences, while also considering quotas and analyzing various scenarios. By providing organizations with informed decision-making capabilities, this system has the potential to be seamlessly integrated into existing information systems. Ultimately, this research contributes towards tackling the intricate challenges associated with refugee resettlement, with promising societal implications.

2 - Shifts in Weather Patterns: A Comprehensive Study of Temperature, Precipitation, and Extreme Events

Alexandros Paparas, Eastern Washington University, Spokane, WA, United States, Stella Alevra, S.B. Fotopoulos

The study examines shifts in climatic variables—temperature, precipitation, and extreme events—across 120 stations in the contiguous United States. Our primary goal is to understand climatic changes and their impact on human life and ecosystems. Through multivariate change point analysis, we identify significant shifts in weather patterns. By adopting this multivariate approach, we gain nuanced insights into dependencies among cities, offering a better understanding of regional climatic variations across diverse geographical landscapes. Our primary method for estimating and inferring change points involves an iterative procedure that combines squared loss-based change point recovery with ℓ_1-regularized squared loss recovery of mean estimates. Although initially developed for single change points, we extend our method to handle multiple change points using the principle of binary segmentation. Our dataset spans from 1948 to 2023, providing comprehensive temporal and geographical coverage. It includes major metropolitan areas and rural surroundings, offering a holistic representation of climatic conditions. Understanding these patterns is crucial for adaptation, timely interventions, and mitigating extreme climatic events.

TD20

Summit - 340

Integrating Emerging Technologies for Sustainable Transportation

Invited Session

Decision Analysis Society

Chair: Xi Cheng, University of Illinois Chicago, Fort Worth, TX, 60607, United States

1 - The Effect of Urban Infrastructure Change on Movement: Investigating Impacts on Visitor Patterns, Business Activities, and Traffic in Seattle after the COVID-19 Pandemic

Grace Jia, University of Washington, Seattle, WA, United States, Ekin Ugurel, Steve Mooney, Jon Froehlich, Cynthia Chen

The COVID-19 pandemic and work-from-home policies disrupted retail businesses and led to deserted downtowns. However, this crisis also spurred municipalities and businesses to innovate their urban infrastructure, aiming to adapt to the new normal and encourage pedestrian activity in public spaces. This research aims to investigate the effects of these urban infrastructure innovations on visitor patterns, business activities, and traffic patterns in Seattle, a city that implemented various changes during the pandemic. Leveraging longitudinal data sources including app-based GPS data, Google Street View, and business data, our study quantifies changes at a block-level resolution. We develop novel algorithms to clean and explore these data, addressing issues such as biases and sparsity. Furthermore, we develop methods to quantify changes in urban infrastructure and assess their impact on the region's transportation system. Our study site, the City of Ballard, is a vibrant business district and has seen significant changes during the COVID period. Additionally, we will identify additional sites in the region to capture a diverse range of built environment characteristics. By answering critical questions about the effects of urban infrastructure innovations, our research will provide valuable policy guidance to local jurisdictions.

2 - Enhancing Urban Rail Transit Planning with a Vec-to-Seq Neural Network

Zhuo Han, University of Massachusetts Amherst, Amherst, MA, United States, Eleni Christofa, Eric Gonzales, Jimi Oke

Urban rail transit (URT) systems are crucial for improving the mobility of urban areas, yet they face challenges in effective planning and operation. This research represents an approach to planning train operations, specifically the Massachusetts Bay Transportation Authority (MBTA) urban rail network in Boston. We estimated a vector-to-sequence (vec-to-seq) neural network aimed at improving URT planning by mapping high-level planning metrics to low-level train operating variables. We used a transformer neural network to evaluate the efficient mapping of inputs to outputs. This study offers a generative model trained to map high-level metrics to movement variables, proving its effectiveness in driving energy efficiency.

3 - Optimizing Two-Echelon Autonomous Modular Vehicle Ridesharing and Consolidation Using Reinforcement Learning

Xi Cheng, University of Illinois Chicago, Chicago, IL, United States, Amir Shafiee, Hanieh Rastegar, Jane Lin

This study introduces a novel approach to optimizing two-echelon ridesharing systems through reinforcement learning with autonomous modular vehicle technology (AMVT). In the first echelon, AMVT pods within suburban areas engage in ridesharing, efficiently collecting passengers who have made requests via a mobile app. In the second echelon, occupied pods consolidate into a pod train, minimizing the number of vehicles and optimizing passenger distribution. Unoccupied pods disjoin from the train to continue individual operations, ensuring flexibility and responsiveness to new requests. To prevent congestion near the on-ramp leading to the downtown expressway, the consolidation point is strategically established based on anticipated demand for subsequent passenger pickups. This dynamic and adaptive approach enhances operational efficiency, reduces congestion, and improves passenger experience in urban transportation systems.

4 - Using Autonomous and Modular Vehicle Technology for Integrated Passenger and Freight Transportation

Amir Shafiee, University of Illinois Chicago, Chicago, IL, United States, Xi Cheng, Hanieh Rastegar, Jane Lin

The surge in urban mobility demands, driven by increasing e-commerce and urbanization, requires more efficient and reliable transportation solutions. We proposed a system that employs Autonomous and Modular Vehicle Technology (AMVT) and explore the advantages of passenger and freight co-modality while evaluating the flexibility of modular adjustments. The vehicle’s modules are designed to serve either freight or passengers and can switch roles depending on current demands. Modules within each section are connected, enabling passengers or cargo to transfer between them during the journey, and also reduce energy consumption resulting from platooning. For security, the connection between the passenger and freight sections can be locked. At the start of the day, the total demand for both passenger and freight transport is known, including preferred pickup and delivery time windows and locations. Vehicles depart from a depot in a pre-determined formation and have the capability to alter their configuration at various intersections. The primary objective of this study is to minimize vehicle operating costs by optimizing the route of vehicles, configuring vehicle formations at each node, and efficiently allocating freight. We also aim to identify the most important factors influencing system performance. Through a mixed-integer programming model, we demonstrate how the integration of modular vehicles and co-modality can significantly enhance transportation efficiency and decrease total operation cost compared to traditional systems.

TD21

Summit - 341

Generative AI for Decision-Making

Invited Session

Decision Analysis Society

Chair: Guanting Chen, UNC Chapel Hill, Chapel Hill,, NC, United States

Co-Chair: Xiaocheng Li, Imperial College Business School, London, United Kingdom

1 - Can LLMs Augment Online Experimentation? Evidence from Headline Selection

Yufeng Zheng, University of Toronto, Toronto, ON, Canada

2 - ORLM: Training Large Language Models for Optimization Modeling

Zizhuo Wang, Chinese University of Hong Kong, Shenzhen, Hong Kong, Hong Kong

Large Language Models (LLMs) have emerged as powerful tools for tackling complex Operations Research (OR) problem by providing the capacity in automating optimization modeling. However, current methodologies heavily rely on prompt engineering (e.g., multi-agent cooperation) with proprietary LLMs, raising data privacy concerns that could be prohibitive in industry applications. To tackle this issue, we propose training open-source LLMs for optimization modeling. We identify four critical requirements for the training dataset of OR LLMs, design and implement OR-INSTRUCT, a semi-automated process for creating synthetic data tailored to specific requirements. We also introduce the IndustryOR benchmark, the first industrial benchmark for testing LLMs on solving real-world OR problems. We apply the data from OR-INSTRUCT to various open-source LLMs of 7b size (termed as ORLMs), resulting in a significantly improved capability for optimiza- tion modeling. Our best-performing ORLM achieves state-of-the-art performance on the NL4OPT, MAMO, and IndustryOR benchmarks.

3 - LLM for the Supply Chain Beer Game

Ming Fan, University of Washington, Seattle, WA, United States, Apurva Jain, Léonard Boussioux, Hongyu Chen

We present a novel approach to designing agentic systems for complex decision-making under uncertainty, using the Beer Game supply chain simulation as a testbed. While state-of-the-art large language models (LLMs) possess advanced human language abilities, they still lack domain-specific knowledge and reasoning capabilities in specialized areas. We explore enhancing ChatGPT's advanced reasoning capabilities to make rational decisions in the context of the bullwhip effect, characterized by demand variability amplification along the supply chain. A Socratic iterative prompting approach is used to iteratively refine prompts and address the deficiency of LLMs in complex decision-making.

4 - Can large language models provide useful feedback on research papers? A large-scale empirical analysis

Hancheng Cao, Stanford University, Stanford, CA, United States

The rapid growth of scholarly production and specialized knowledge challenges conventional scientific feedback mechanisms, making high-quality peer reviews increasingly difficult to obtain. With the advent of large language models (LLMs) like GPT-4, there's growing interest in using them for generating scientific feedback on research manuscripts. However, the utility of LLM-generated feedback has not been systematically studied. To address this, we developed an automated pipeline using GPT-4 to provide comments on full PDFs of scientific papers. We evaluated the quality of GPT-4's feedback through two large-scale studies. First, we quantitatively compared GPT-4's feedback with human peer reviewers' feedback across 15 Nature family journals (3,096 papers) and the ICLR machine learning conference (1,709 papers). The overlap in points raised by GPT-4 and human reviewers (30.85% for Nature journals, 39.23% for ICLR) is comparable to the overlap between two human reviewers (28.58% for Nature journals, 35.25% for ICLR). For weaker papers (rejected ICLR papers), GPT-4's overlap with human reviewers was higher (43.80%). In a prospective user study with 308 researchers from 110 US institutions in AI and computational biology, 57.4% found GPT-4's feedback helpful/very helpful, and 82.4% found it more beneficial than feedback from some human reviewers. Despite its benefits, GPT-4 has limitations, such as focusing on specific feedback aspects and struggling with in-depth critique of method design. Our results suggest that while human expert review remains essential, LLM feedback can complement it, especially when timely expert feedback is unavailable and during early manuscript preparation stages.

TD22

Summit - 342

Behavioral Decision Analysis

Invited Session

Decision Analysis Society

Chair: Yucheng Su, IE Business School, MADRID, 28006, Spain

Co-Chair: Matthias Seifert, Operations & Business Analytics, IE Business School - IE University, Madrid, Spain

1 - Introducing the Significance Fadeaway Score

Pranadharthiharan Narayanan, Universidade Nova de Lisboa – Nova School of Business and Economics. VAT: PT501559094, Lisbon, Portugal, Sumit Malik

We introduce a simple metric called "significance fadeaway score" (SFS) to examine the robustness of t-test results in experimental studies. SFS assesses the extent to which statistical significance, typically indicated by $p < 0.05$ hinges on extreme values. To compute SFS, we iteratively remove the most extreme data points (one-at-a-time and without replacement) and recalculate p-values. The resulting SFS metric is highly interpretable. For instance, an SFS of 7 means that statistical significance would fade away if we exclude 7% of most extreme data points. We empirically evaluate SFS scores from ~50 published experimental studies in leading marketing journals (JCR and JCP) and find nearly half the studies lose statistically significance if we selectively remove less than 2% of the sample. We believe this result signals a critical reflection point on revaluating our reliance on statistical significance. Notably, p-values are highly sensitive to selective exclusion of small samples, even in the absence of overt outliers.

2 - System Neglect in Newsvendor Decision Making

Yucheng Su, Operations & Business Analytics, IE Business School - IE University, MADRID, Spain, Matthias Seifert, Canan Ulu, SREYAA GUHA

We study the role of system neglect in newsvendor decision making when facing dynamically changing demand conditions. Previous research has shown that probability judgments in response to observing signals of change often suffer from the system neglect bias, in that decision-makers primarily react to the signal itself rather than the environmental system that generated it. In the present paper, we study how system neglect interacts with common biases in inventory decision making, where managers judge probabilities regarding an impending demand shift. We develop an analytic model that predicts empirical newsvendor behavior and compare it to that of a Bayesian risk-neutral agent. We then test our predictions in an experimental study.

3 - How do Newsvendors react to near misses?

Florian Federspiel, INCAE Business School, Escazu, Costa Rica, Robin Dillon-Merrill, Matthias Seifert

We study *when* and *why* repeated near misses in the context of inventory stockouts lead to inappropriate order quantities and risk-taking. Little is known about the behavioral consequences and risks from near misses in inventory decisions. We offer a theoretical foundation and empirical support from several experimental studies to shed light upon a novel bias applicable to newsvendor decision-making, showing that if prior events do not cause the decision maker to reflect on how close a near miss situation could have been to a negative consequence, these repeated events can lead to irrational order quantities increasing one's exposure to stockouts. We demonstrate that as a newsvendor experiences multiple near misses over time without also experiencing stockouts, they will order quantities that increase the risk of exposure to future stockouts. We demonstrate that decision makers in newsvendor contexts are predictably irrational in their reaction to repeated near misses affecting ordering choices going forward.

4 - Incentive-Compatible and Strongly Fair Cake Cutting

Jens Witkowski, Frankfurt School of Finance & Management, Frankfurt, Germany, Rupert Freeman

The classical cake cutting setting is concerned with dividing a resource, modeled by the [0,1] interval, and allocating subintervals to different agents. Preferences of agents are represented by a function over the [0,1] interval with the agent's utility corresponding to the area under the curve for those subintervals they get allocated. A recent result shows that there does not exist a deterministic cake cutting mechanism that is both incentive compatible and even only one of proportional or envy-free (the latter is restricted to non-wasteful mechanisms). In this work, leveraging proper scoring rules, we design a randomized cake cutting mechanism that is ex ante incentive compatible, ex post strongly proportional, and ex post strongly envy-free. This result is tight in the sense that additionally achieving ex post incentive compatibility is impossible as it would violate the aforementioned impossibility result.

5 - Preferences for Wealth Distributions in the Presence of Group Structure

Rupert Freeman, University of Virginia, Charlottesville, VA, United States, Stefano Balietti

We examine individual preferences for wealth inequality in a population that is endowed with group structure. We conduct an online experiment in which we assume that each individual in the population is associated with a single group identity. Using a custom Distribution Builder tool, participants are asked to build their preferred distribution of wealth for various group structures, including the case of separated and well-mixed groups. In the former case, all members of one group are disadvantaged relative to the other, and in the latter, the distribution of wealth is roughly equal across each group. Our results indicate that preferences for inequality differ significantly when group structures are present relative to a baseline case with no group fairness considerations. Our work is relevant to any setting where group fairness considerations are relevant (e.g., hiring, civic participation, public policy, college admissions), and stands in contrast to existing literature on preferences for inequality that does not simultaneously consider inequality at both the individual and group levels.

TD23

Summit - 343

Global Supply Chains

Contributed Session

Chair: Zhaocheng Zhang, University of Cambridge, Cambridge, N/A, United Kingdom

1 - Designing a Stochastic Supply Chain Network: An Error-based Heuristic

Gang Wang, University of Massachusetts Dartmouth, North Dartmouth, MA, United States

Designing a stochastic supply chain is challenging because demand is often partially observable or unknown in advance. This paper presents an alternative approach to address a three-echelon stochastic supply chain network design with insufficient demand information. We formulate a stochastic mixed-integer program with chance constraints. We identify a lower and upper bound on the value at risk to deal with chance constraints, thus inducing two mixed-integer linear approximations to the stochastic program. We develop a branch-and-price scheme to solve the approximations and a primal heuristic algorithm for initial columns. We then create a heuristic algorithm for solving the original problem by implementing sensitivity analysis and error bound on the approximations. Besides network size, computational experiments find that network design performance depends on aggregate service levels and maximum demand variations. The aggregate behavior of service levels has an inverse impact on the network design performance, while maximum coefficients of demand variations present a positive effect. Computational results reveal that our proposed algorithms perform well compared to normal distributions and a robust optimization formulation. Our models and algorithms suggest a feasible network design tool for handling uncertainties in supply chains.

2 - A Variational Inequality Trade Network Model in Prices and Quantities Under Commodity Losses

Samirasadat Samadi, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Ismael Mohammad Pour

Multicommodity trade enables the production, consumption, and flow of commodities across the globe from agricultural ones to precious metals. Mathematical formalisms to model, analyze, and solve such problems have advanced and are also relevant to policy and decision-making. In this paper, we construct a variational inequality trade network model in price and quantity variables, which captures possible losses on transportation routes, which can occur because of perishability of commodities, as in the case of agricultural ones, or outright thefts. The equilibrium conditions are stated and the variational inequality formulation derived. Qualitative properties of existence and uniqueness of the equilibrium supply price, commodity shipment, and demand price pattern are provided under reasonable conditions. Illustrative examples help to demonstrate the model. An algorithm that is proposed yields closed form expressions at each iteration and can also be interpreted as a discrete time adjustment process for the evolution of the economic variables. A spectrum of algorithmically solved numerical examples, with full input and output data provided, yields insights into the impacts of commodity losses, increased congestion, as well as enhanced marketing on producers as well as consumers. This new model expands the scope of spatial price equilibrium modeling under commodity losses.

3 - Tensor Factor Analysis of Global Supply Chains

Zhaocheng Zhang, University of Cambridge, Cambridge, United Kingdom, Weichen Wang, Jing Wu

This paper presents a novel empirical framework for analyzing granular supply chain data. Rather than focusing on supply chain diads, we model the entire global supply chain flows as a higher-order tensor, allowing us to explore the underlying latent lower-dimensional structure of the dynamic supply chain network using recently developed tensor-valued factor models. By examining the latent networks and their connections to the surface networks, we gain a clear view of how supply chains have evolved over the past two decades. The resulting lower-dimensional representation of the dynamic network can then be utilized for second-step analyses, such as predicting the time series of the tensor-valued supply chain network. In addition to the standard factor model, we further introduce characteristic-based factor loadings modeled as unknown functions of country-level and industry-level time-varying covariates. We propose an efficient estimation algorithm for this semiparametric model. The latent lower-dimensional tensor factors estimated from our models capture over 85% of the evolution of global supply chain networks.

TD24

Summit - 344

Adaptive Strategies in Experimentation and Logistics

Contributed Session

Chair: Yifan Wu, 2250 Dartmouth St, 77840

1 - On the Value of Flexibility for Adaptive Experiments

Huijun Zhu, National University of Singapore, Singapore, Singapore, Yifan Feng, Tu Ni

Adaptive experimentation refers to the practice of changing (improving) the experiment configurations on the fly based on observed data. While it can potentially increase sample efficiency, it potentially needs strong flexibility as an experimenter and thus may be difficult to implement in practice. To explore the interplay between sample efficiency and flexibility in adaptive experiments, we formulate a Bayesian sequential hypothesis testing problem. The goal is to minimize the sum of sampling and penalty costs under different levels of flexibility. Here, flexibility is defined as the number of times one can change the experiments. We show that there is a notable gap between "No Flexibility" (where experiments cannot be changed) and "Full Flexibility" (where experiments can be freely changed adaptively). Nevertheless, the gap can be (asymptotically) closed by just allowing to change the experiment once. That is achieved by an experimentation rule we develop, termed Nested SPRT, which is both simple to derive in closed form and easy to implement.

2 - A vehicle routing problem for massive Covid-19 testing in Antofagasta, Chile

Hernan Caceres, Universidad Catolica del Norte, Antofagasta, Chile, Javiera Auad, Andrea Fernández, Blanca Peñaloza, María-Soledad Zuzulich

The COVID-19 pandemic has demanded massive and recurrent testing due to the high number of cases that have been found in the last two years. Mass testing requires significant resource management to develop, which is why a useful tool to manage them is optimization. A project was developed in Antofagasta, Chile, to help small businesses to have more control over positive cases in their establishments. A medical team would visit them regularly to have greater control and provide them with more security. To do that efficiently, a vehicle routing problem with multiple time windows was developed to optimize the time of visit routes and medical equipment necessary to carry out the testing.

3 - Discrete Event Simulation inside a Private Hospital in Brazil

Reinaldo Crispiniano Garcia, Universidade de Brasilia, Brasilia, Brazil, Joao Gabriel De Souza Vale, Ricardo Accorsi, Antonildes Nascimento Assuncao, Noerci Batistela Junior

The COVID-19 pandemic has had a profound impact on global health systems, leading to unprecedented demand for emergency services and then the need for innovative operational solutions. This research aimed to design a Discrete Event Simulation (DES) model of an Emergency Department (ED) in a Brazilian private hospital to identify process bottlenecks. The studied hospital is one of the main Brazilian Hospitals attending more than 40 thousand patients yearly. The analyzed process had different specialities including pediatrics, orthopedics, geriatrics and neurology ones. Moreover, not only the different arrival rates for each medical practice was considered but also their seasonality and their service rates. It was then observed that just scheduling the physicians according to the patients bottlenecks, it was possible to reduce the patients total time in the system by about 5%. Once the efficiency increase was obtained without raising the systems cost, the applied approach is a promising one to increase the hospitals efficiency in a time of scarce health resources.

4 - Train, then Estimate and Decide

Tarkan Temizoz, Eindhoven University of Technology, Eindhoven, Netherlands, Christina Imdahl, Remco Dijkman, Douniel Lamghari-Idrissi, Willem van Jaarsveld

The effectiveness of data-driven strategies in inventory management is undergoing reevaluation. The COVID-19 pandemic highlighted the limitations of relying solely on historical data as industries worldwide grappled with drastic changes in demand patterns and supply uncertainties. Additionally, there can be scenarios with limited or no historical data, such as new product launches, where data is concurrently acquired with inventory decisions. These challenges are particularly compounded when data is censored, with decision-makers only able to observe sales but not the actual demand or not knowing when the placed orders will come due to supply uncertainty.

Addressing limited, censored data necessitates modeling inventory problems as Partially Observable Markov Decision Processes. In this work, we propose a new framework to solve POMDPs, typically encountered in inventory management when the demand or the lead time distribution is censored. This framework is based on Train, then Estimate and Decide strategy. Specifically, for solving POMDPs corresponding to a generic inventory model, we train a neural network policy over the possible parameter spaces of the unknown or censored variables using deep reinforcement learning. Then, upon collecting possibly censored data with new observations, we estimate the related parameters and feed the trained neural network with them for replenishment decisions.

Experimental results reveal that our framework, equipped with a state-of-the-art DRL algorithm -Deep Controlled Learning- for the training task and Kaplan-Meier Estimator for the estimation task, outperforms the best online learning algorithms in lost sales inventory control. This level of performance is attained using a single neural network policy.

5 - Revisiting Continuous p-Hub Location Problems with the L1 Metric

Yifan Wu, Texas A&M University, College Station, TX, United States, Joseph Geunes, Xiaofeng Nie

Motivated by emerging urban applications in commercial, public-sector, and humanitarian logistics, we revisit continuous p-hub location problems under the L1 metric, where several facilities are to be located in a continuous space such that the expected minimum Manhattan travel distance from a random service provider to a random receiver through exactly one hub facility is minimized. Firstly, some closed-form results are derived for one-dimensional cases and two-dimensional cases up to two hubs. Secondly, we propose a simulation-based approximation method for problems with more than two hubs in two-dimensional cases. Finally, we apply our method to a case study on optimizing public access defibrillator deployment.

TD25

Summit - 345

CoCalc/Amazon

Invited Session

Technology Showcase

1 - Using AI in CoCalc to Improve your Collaborative Workflow

Blaec Bejarano, CoCalc by SageMath, Inc., Renton, WA, United States

Explore how CoCalc can enhance your collaborative projects at the INFORMS 2024 Annual Meeting. Join us to see how AI-driven features within CoCalc can streamline tasks, improve data analysis, and facilitate teamwork for computational scientists. This session offers practical insights and strategies you can immediately apply to make your collaborative efforts more effective. Don't miss this opportunity to see how CoCalc can help you reduce friction between different workflows within your organization.

The content presented is most relevant to Professional (Mid-Career); Executive (Senior Level); and Associate (Early Career).

2 - Constraints and Coordination for Deep Reinforcement Learning Agents

Carson Eisenach, Amazon, Inc, New York, NY, United States

We explore new techniques for constrained reinforcement learning (RL) in the real-world, specifically as applied to inventory management. The classic approach is to use model predictive control to enforce constraint adherence. With Deep RL policies, this becomes complicated as they consume high-dimensional features to make decisions, and accurately forward simulating this joint distribution is extremely challenging, if not impossible. This session provides an overview of a new approach -- "Neural Coordinator" -- which directly forecasts dual costs under which a policy will adhere to the desired constraints. We will demonstrate its effectiveness in the inventory control setting and cover how to backtest policies in the presence of constraints.

TD26

Summit - 346

Behavioural Queueing Science and Service Operations

Invited Session

MSOM: Service Operations

Chair: Yichuan Ding, McGill University, Montreal, QC, H3A 1G5, Canada

Co-Chair: Yifeng Cao, University of British Columbia, Vancouver, BC, Canada

1 - Dynamic Transfer Policies for Parallel Queues

Jangwon Park, University of Toronto, Toronto, ON, Canada, Timothy Chan, Vahid Sarhangian

We consider the problem of load balancing in parallel queues by transferring customers between them at discrete points in time. Holding costs accrue as customers wait in the queue, while transfer decisions incur both fixed (setup) and variable costs proportional to the number and direction of transfers. Our work is primarily motivated by inter-facility patient transfers between hospitals during a surge in demand for hospitalization (e.g., during a pandemic). By analyzing an associated fluid control problem, we show that under fairly general assumptions including time-varying arrivals and convex increasing holding costs, the optimal policy in each period partitions the state-space into a well-defined *no-transfer region* and its complement, such that transferring is optimal if and only if the system is sufficiently imbalanced. In the absence of fixed transfer costs, an optimal policy moves the state to the no-transfer region's boundary; in contrast, with fixed costs, the state is moved to the no-transfer region's relative interior. We further leverage the fluid control problem to provide insights on the trade-off between holding and transfer costs, emphasizing the importance of preventing excessive idleness when transfers are not feasible in continuous-time. Using simulation experiments, we investigate the performance and robustness of the fluid policy for the stochastic system. In particular, our case study calibrated using data during the pandemic in the Greater Toronto Area demonstrates that transferring patients between hospitals could result in up to 27.7% reduction in total cost with relatively few transfers.

2 - Foresee the Next Line: Customer Strategies and Information Disclosure in Tandem Queues

Ricky Roet-Green, Simon Business School, University of Rochester, Rochester, NY, United States, Jingwei Ji, Ran Snitkovsky

Many services consist of multiple stages, where each stage requires some waiting before completion. For example, customers who visit the Apple Store join the check-in queue first, and then wait in another queue to be served by the Genius Bar technician. Although customers may observe the queue in front of them, they usually have no information about the waiting situation in the next queue. Our paper aims to examine the impact of queue-length information on customers' strategic behavior in such systems. We assume a two-stage tandem queueing system, with an admission queue followed by a treatment queue. Customers observe the queue length at arrival to each queue; they may balk or join and might later renege. We first study the fully observable model, in which queue-length information of both queues is available to customers at the time they arrive to the system. We calculate the equilibrium strategy and show that it is not necessarily a function of the total number of customers in the system. Next, we study the partially observable model, in which customers observe each queue length only at arrival to it, i.e., they do not observe the second queue length when they arrive to the system. We compare the system performance across the fully and partially observable models. We find that in most cases the partially observable model yields higher throughput but lower social welfare compared to the fully observable model.

3 - Should Multi-Listing be Allowed? Throughput and Social Welfare Analysis

Yichuan Ding, McGill University, Montreal, QC, Canada

We consider a system that consists of multiple service providers (SP), each of which keeps its own waitlist. The service offered by different SPs have similar content and thus a customer, despite having preference between different SPs, can be served by any one of the SPs and leave the system. In this case, some customers may opt to register on multiple waitlists for early access to service. We refer to such queue-joining behavior as “multi-listing”, which is ubiquitous in real life. We analyze the throughput rate and the social welfare in a multi-listing system and derive insights into the management of such a system. We also compare it with other system designs such as a pooling queue or a separate-queue system.

TD27

Summit - 347

Exploring Challenges for Innovation Success

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Param Pal Singh Chhabra, University of Alberta, Edmonton, AB, Canada

Co-Chair: Gulru Ozkan-Seely, University of Washington, Bothell, WA, United States

1 - Impact of Successful Drug Launches on Biotech-Pharma Alliance Terminations: A Machine Learning Approach

Pierre Gautreau, York University - Schulich School of Business, Toronto, ON, Canada, Moren Levesque, Annapoornima Subramanian, Vareska van de Vrande

This study investigates the phenomenon of biotech-pharma alliance breakups following successful drug launches by pharmaceutical partners, a vulnerability to young biotech firms unexplored in the current literature. Using a data-driven theory-building approach, we employ machine learning techniques to analyze 3,717 terminated biopharmaceutical alliances from 2000 to 2019, sourced from the Cortellis Drug Discovery database. Our methodology relies on Bayesian Additive Regression Trees (BART) to isolate alliance terminations induced by the pharmaceutical partner’s successful drug launch and evaluates the conditions more likely to yield such termination. Our findings reveal that a successful drug launch by a pharmaceutical partner increases the likelihood of alliance termination by 6%. Notably, pharmaceutical firms with high R&D intensity are less likely to terminate alliances post-launch, indicating the importance of finding a partner that is actively engaged in R&D. Further analysis identifies firm size, failure experience, and the mode of alliance formation (equity versus non-equity alliance) as significant determinants of alliance vulnerability to a pharmaceutical partner's drug launch. The study contributes to the literature on innovation alliances and R&D productivity by demonstrating the premature nature of these terminations, which adversely affect both biotech firms' operational capabilities and the pharmaceutical firms' R&D productivity. Additionally, this research underscores the importance of strategic considerations in alliance formation to mitigate the risks associated with successful drug launches. Our work also provide practical insights for manager to consider when selecting alliance partners.

2 - Patent Pendency and Future Innovative Activities

PARAM PAL SINGH CHHABRA, University of Alberta, Edmonton, AB, Canada, Manpreet Hora, Karthik Ramachandran

We study the effect of patent pendency on inventors’ future innovative activities. We find support for our hypotheses by testing them using inventor-level patent data published by the United States and Patent Trademark Office (USPTO). Our results indicate a negative effect of longer patent pendency on the number and quality of inventors' future patents.

Keywords: patent pendency, inventors, resource allocation, USPTO.

3 - Role of Startup Rating Agencies in Improving Venture Investment Decisions

Thunyarat Amornpetchkul, Santa Clara University, Santa Clara, CA, United States

The emergence of startup rating agencies in the contemporary startup ecosystem has sparked interest but remains largely unexplored in academia. These agencies offer rigorous evaluations and due diligence on ventures, potentially enhancing insights into their success prospects. However, their services typically come at a cost, either in the form of a fixed fee or an equity share. This study aims to identify the circumstances under which an investor finds it beneficial to engage startup rating agencies to obtain more accurate venture information for informed investment decisions. We analyze the investor’s optimal decisions under two payment schemes: a fixed fee scheme and an equity share scheme, and subsequently determine the conditions under which each scheme is preferable. Additionally, we analyze the problem from an entrepreneur’s perspective to see when it would be worthwhile for her to pay for a startup rating agency’s service herself, expecting to influence the investor’s decisions in her favor. Through these analyses, we aim to provide a comprehensive understanding of the implication and efficacy of startup rating agencies in improving venture investment outcomes.

4 - Emerging-Economy Patent Regulation Reforms and Firm Innovativeness Abroad

Xiaojin Liu, Virginia Commonwealth University, Richmond, VA, United States, Pankaj Kumar, Aks Zaheer

Emerging-economy firms are making significant efforts to move up the value chain by improving their innovativeness and increasing innovations abroad. We investigate whether and under what conditions patent regulation reforms in emerging economies affect these firms’ outward innovation behavior. Our result highlights that emerging-economy firms become more internationally competitive post-reform.

5 - A Machine Learning Approach to Predicting Project Performance

siqi wang, University of California, San Diego, La Jolla, CA, United States, Xiaochen Gao, Sina Khorasani, Vish Krishnan, Lakshminarayana Nittala

We use a massive U.S. government dataset to predict cost overrun and schedule delay in government contracts. Despite facing a high-dimensional dataset, we manage to build a well-performing CatBoost model with exceptionally high prediction accuracy. We identify top predictors for cost and schedule overruns, interpret our predictions using SHAP values, and offer recommendation based on causal models that could lead to substantial time and cost savings.

TD28

Summit - 348

Optimizing Data Center Operations and Sustainability

Contributed Session

1 - Data Value in Distribution System Operations

Mehrnoush Ghazanfariharandi, Rutgers University, Piscataway, NJ, United States, Robert Mieth

The rise of advanced data technologies in electric power distribution systems enables operators to optimize operations, but raises concerns about data security and consumer privacy. Resulting protection mechanisms that alter or obfuscate datasets may invalidate the efficacy of data-driven decision-support tools and impact the value of these datasets. This paper derives tools for distribution system operators to enrich data-driven operative decisions with information on data quality and, simultaneously, assess data usefulness in the context of this decision. To this end, we derive an AC optimal power flow model for radial distribution systems with data-informed stochastic parameters that internalizes a data-quality metric. We derive a tractable reformulation and discuss the marginal sensitivity of the optimal solution as a proxy for data value. Our model can capture clustered data provision, e.g., from resource aggregators, and internalize individual data quality information from each data provider. We use the IEEE 33-bus test system, examining scenarios with varying photovoltaic penetration, to demonstrate the application of our approach and discuss the relationship between data quality and its value.

2 - Momentum Power Forecast: A Statistical Benchmark to Improve Google's Data Center Capacity Planning

Lijuan Xu, Google, Mountain View, CA, United States

Google builds a global footprint of data centers to support our products and services. Accurate and cross functionally aligned demand forecasts are essential to inform the right data center capacity strategies and coordinate the various planning activities. Over decades of learning and practice, we developed the "human-in-the-loop" forecasting framework to integrate data science with business planning. In this talk, we are happy to share the behind-the-scenes cooking of our statistical benchmark power forecasts that bring forecasting, metrics, and engineering automation all together to influence Google's business planning in the data center space.

3 - Predicting Potential Uncorrectable Errors Using Telemetry AI Modeling for Data Center Reliability

Youngmin Lee, Samsung Electronics, Hwaseong, Korea, Republic of

In data centers, uncorrectable errors (UEs) occurring in DRAM can potentially render servers inoperative, making it a critical issue to reduce such errors. We introduce a novel UE Detective methodology to predict risky DRAMs based on data center's telemetry logs, which contain detailed information about correctable errors (CEs). First, we develop spatio-temporal features to help predict failures by utilizing domain information based on the actual UE judgment mechanism from the bit-level information of CEs. Furthermore, we implement Positive-Unlabeled (PU) learning to enhance our predictive model's robustness in managing the label uncertainty of normal DRAMs that, due to the stochastic nature of server operations, display UE symptoms in telemetry logs but have not yet manifested UEs. Subsequently, we apply cutting-edge deep learning models for tabular data to predict UEs, exhibiting superior performance compared to traditional tree models. Our proposed methodologies have been empirically validated through long-term real-world experiments using extensive telemetry data from a data center customer, showcasing the potential to preemptively identify over 70% of DRAMs encountering UEs. Our methodology can trigger a pullout process that proactively remove potentially faulty DRAMs, preventing actual UEs from manifesting and thus contributing to the reliability of the server.

4 - Reinforcement Learning-based scheduling of PV panel cleaning in arid regions considering uncertain dusty weather conditions

Heungjo An, Kumoh National Institute of Technology, Gumi, Korea, Republic of, Hyunsoo Lee

Airborne dust easily accumulates on the surfaces of solar panels, diminishing power output particularly in arid regions. A common strategy to mitigate dust deposition involves periodic cleaning of photovoltaic (PV) panels. However, the frequency of cleaning impacts the economic viability of solar PV systems, presenting a trade-off between cleaning costs and energy loss. To tackle this challenge, this study explores various metrics and devises a generic framework based on Reinforcement Learning to dynamically determine the cleaning intervals for PV panels, accounting for uncertain dusty weather conditions. Additionally, a simulation and optimization model is developed to establish a constant cleaning interval for the entire planning horizon (i.e., 20 years) for comparative analysis. Finally, this study contrasts the economic and energy-related metrics derived from the proposed RL-based method with those from the simulation optimization approach.

5 - Is ESG "Worth" Anything? On the Economic Value of Climate Pledges on Amazon

Nuo Yuan, Boston University, Boston, MA, United States, Yuze Li, Dokyun Lee

In response to the growing consumer demand for sustainability, many firms have adopted product certifications issued by third-party agencies to signal their products' social and environmental sustainability. In this study, we make the first empirical attempt to examine the economic impact of these sustainable product certifications. Using a unique cross-sectional dataset of facial skincare products sold on Amazon, we employ a mixed-method approach that combines machine learning and causal inference to examine the causal effect of "Climate-Pledge" certification on product sales rank in a cross-category fashion, while accounting for other product attributes identified by prior literature as drivers of sales rank (e.g., reviews and Q&As). We also conduct a series of moderation analyses based on further segmentations by product category and certificate type. Focusing on climate-pledge product certificates that have emerged due to increasing demands for product transparency from "green" consumers, the results of this study hold important implications for Amazon's partnering manufacturers and the platform itself.

TD29

Summit - 420

Warm Starting for Mixed-Integer Linear Programming

Invited Session

OPT: Integer and Discrete Optimization

Chair: Shannon Kelley, Lehigh University, 337 South New Street, Chicago, IL, 18015, United States

1 - Warm Starting of Mixed Integer Linear Optimization Problems via Parametric Disjunctive Cuts

Shannon Kelley, Lehigh University, Bethlehem, PA, United States, Aleksandr Kazachkov, Ted Ralphs

Many applications require solving a series of closely-related mixed integer linear optimization problems that all share the same variables and number of constraints. In such cases, warm starting — reusing information from a previous solve — provides an opportunity to improve overall solution time. In this paper, we explore the use of valid disjunctions generated from branch-and-bound trees arising during previous solves, along with so-called Farkas certificates that prove the validity of associated disjunctive cuts, to warm start the solution of instances later in the sequence. The combination of a disjunction and a Farkas certificate yields a parametric class of disjunctive cuts that can be inexpensively applied to any instance in the sequence. We implemented this idea and performed computation experiments to test its potential.

2 - Mixed integer branch and bound reoptimization for parametric problems

Stefan Clarke, Princeton University, Princeton, NJ, United States, Gabriele Dragotto, Bartolomeo Stellato

We study a parametric family of mixed integer linear programs. We present a model which creates a branch-and-bound tree in an offline phase, and then rapidly searches over it in an online phase to ensure fast online optimization.

3 - A Cycle-Cancelling Heuristic for The Traveling Salesman Problem

Steffen Borgwardt, University of Colorado-Denver, Denver, CO, United States, Zachary Sorenson

The Traveling Salesman Problem (TSP) is one of the most-studied hard problems in combinatorial optimization. We developed a new algorithm that uses a connection between Minimum Cost Flow Problems (MCF) and the TSP to improve on a given local optimum.

MCF problems can be solved efficiently through linear programming or combinatorial algorithms such as cycle-cancelling. We investigated the potential of cycle-cancelling in the context of the TSP. Through a restriction of the search space of cycles to cancel for a given tour, practical results exhibit that only a low number of subtours is created, and a simple patching step suffices for a high success rate and gap closure towards an optimum.

4 - New cutting planes for two-stage stochastic integer programming: neighbor cut

jingye xu, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey, Diego Cifuentes

cutting plane methods are the most popular method for solving two-stage stochastic integer programs where one iteratively generates cuts to approximate the second-stage value function. Those cuts are usually generated from certain optimal dual variables of some convex programs. One can also view cutting plane methods as iteratively conducting sensitivity analysis on the second-stage value function. However, from the perspective of sensitivity analysis, optimal dual variables provide good predictions only if the perturbation is small. This contradicts the fact that in many two-stage stochastic integer programs, the distance of any two distinct state variables are large (at least one). This motives us to derive new cutting planes from near optimal dual variables instead of optimal dual variables. Such cuts lead to much better prediction on "neighbors" of any state variables and are very easy to implemented. Results of numerical experiments are presented. In many cases, such new simple cuts leads to speed up solving time by magnitudes.

TD30

Summit - 421

Large Scale Supply Chain Network Design

Invited Session

OPT: Computational Optimization and Software

Chair: Arash Haddadan, Amazon.com, Bellevue, WA, United States

Co-Chair: Chun Ye, Amazon.com, Seattle, WA, United States

1 - Solving the Continuous TIME Service Network Design Problem (Ctsnd) by Column Generation

Marcus Poggi, Amazon, Seattle, WA, United States, Arash Haddadan, Daniel Ulch, Jochen Koenemann, Madison Van Dyk

Given a network and a set of commodities, the CTSND seeks a set of vehicle trips to transport the commodities from their origins to destinations minimizing the total cost. Each commodity has a release time at origin, a due time at destination, and an associated quantity. Transit times, vehicle capacities, fixed cost and cost per unit transported are specified for each arc. A feasible solution describes the set of trips on each arc, their departure times and the commodities they carry. We show that considering at most two times the number of commodities departures times for each arc is enough to determine the continuous time optimal solution. Our approach models the commodities as flows on the untimed network and require the flow of each commodity through each arc to be associate with a vehicle trip. Time constraints assure the commodities start their trip after their release date, depart at from any intermediary node only after its arrival, and completes its journey before its due time. A column generation subproblem is solved for each potential departure time on each arc. Candidate commodities are selected for each potential trip in order to minimize the associated trip reduced cost. We present results for instances from the main CTSND benchmarks and for large instances from the logistics operation of a large web-commerce company. They show the model proposed provide strong lower bounds. The linear relaxation solution is the input for a sequential rounding procedure which produces high quality primal solutions.

2 - Designing Delivery-Anchors for Amazon's Middle Mile Network

Arash Haddadan, Amazon.com, Bellevue, WA, United States, Jochen Koenemann, Leian Chen, Nityansh Seth

Amazon's ability to deliver on estimated shipping speeds and build customer trust hinges on the design and scheduling of its "middle-mile" supply chain network. The objective is to optimize not just costs, but also delivery speed estimates provided to customers. Amazon utilizes "delivery-anchors" to offer concrete estimated delivery dates. Delivery-anchors are deadlines for performing operations in the network, such as transportation along a lane. They act as delivery certificates, allowing Amazon's fulfillment systems to make accurate shipping estimates. Designing delivery-anchors effectively is a complex optimization problem that requires considering both fulfillment software and operational constraints. In this work, we propose a mixed-integer-programming model for delivery-anchor design, building on the Service Network Design problem. The delivery-anchor optimization enables Amazon to cost-effectively design its middle-mile network to achieve faster customer delivery speeds. Our design considers fulfillment engine requirements and real-world operations. Since our model uses a time-expanded network, solving large-scale instances is computationally challenging without heuristics. We employ size-reducing assumptions and model simplifications to enable scalability. Experimental results validate our model's usefulness for optimizing Amazon's delivery-anchor network design against the dual objectives of cost and delivery speed targets..

3 - Designing Optimization Software for Use-Case Scalability

Semih Atakan, Amazon.com, Seattle, WA, United States

The scalability of optimization models is a common concern in academic literature, but in industry, scalability to multiple use cases is equally critical. Often overlooked, this aspect results in a proliferation of optimization models, each with similar mathematical structures, but specialized software for solving distinct problems. This presentation delves into the design of software solutions that enable resolution of diverse business challenges at scale.

4 - Region-Based Network Design

Baris Burnak, Amazon.com, Bellevue, WA, United States, Semih Atakan, Jochen Koenemann

In 2023, Amazon launched Regionalization of the US network, which partitioned the US into smaller and roughly self-contained regions. This work presents a two-stage region-based decomposition algorithm for the fixed-charge multicommodity flow problem by leveraging the Regionalization structure to reduce the search space and improve the computational time to close the optimality gap. The first stage involves solving region-level subnetworks and fixing a subset of region-level decisions. In the second stage, we solve for the full network with fixed regional decisions. The key challenge to decompose the US network is the a priori allocation of the shared resources across subnetworks. Our proposed methodology distributes these resources for the fulfillment of regional and cross-regional commodities. This decomposition algorithm significantly reduces the computational time to close the optimality gap compared to a generalized monolithic algorithm, while producing solutions that are within 1% of the cost of the true global optimum solution.

TD31

Summit - 422

Promises and Challenges of Nonconvex Optimization in Machine Learning

Invited Session

OPT: Nonlinear Optimization

Chair: Salar Fattahi, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Jianhao Ma, University of Michigan, Ann Arbor, MI, United States

1 - Parsimonious Trajectories in Gradient Descent: Emerging Low-dimensionality in Over-parameterized Models

Jianhao Ma, University of Michigan, Ann Arbor, MI, United States, Salar Fattahi, Geyu Liang

Recent studies have shown that Gradient descent (GD) and its variants inherently favor "parsimonious" solutions, tending towards optimal solutions in a low-dimensional region even within overparameterized spaces. This paper explores the underlying mechanics of this phenomenon, establishing minimal requisite conditions under both convex and nonconvex settings. We demonstrate that certain GD variants, under these conditions, efficiently target parsimonious solutions. Intriguingly, we find that Perturbed Gradient Descent (PGD) that reaches second-order stationary points efficiently in the classic nonconvex settings underperforms in standard overparameterized nonconvex problems like matrix factorization. To address this, we introduce the Two-Sided Perturbed Gradient Descent (TSP-GD), an algorithm tailored for complex nonconvex optimization. TSP-GD achieves rapid convergence to second-order stationary points in the overparameterized regime, akin to PGD in low-dimensional spaces. This innovation in TSP-GD could be of independent interest to the nonconvex optimization community.

2 - On the Trajectories of SGD Without Replacement

Pierfrancesco Beneventano, Princeton University, Princeton, NJ, United States

We examine the implicit regularization effect of Stochastic Gradient Descent (SGD). We consider the case of SGD without replacement, the variant typically used to optimize large-scale neural networks. We analyze this algorithm in a more realistic regime than typically considered in theoretical works on SGD, as, e.g., we allow the product of the learning rate and Hessian to be O(1) and we do not specify any model architecture, learning task, or loss (objective) function.
Our core theoretical result is that optimizing with SGD without replacement is locally equivalent to making an additional step on a novel regularizer.
This implies that the expected trajectories of SGD without replacement can be decoupled in (i) following SGD with replacement (in which batches are sampled i.i.d.) along the directions of high curvature, and (ii) regularizing the trace of the noise covariance along the flat ones.
As a consequence, SGD without replacement travels flat areas and may escape saddles significantly faster than SGD with replacement.
On several vision tasks, the novel regularizer penalizes a weighted trace of the Fisher Matrix, thus encouraging sparsity in the spectrum of the Hessian of the loss in line with empirical observations from prior work.
We also propose an explanation for why SGD does not train at the edge of stability (as opposed to GD).

3 - Trained Transformers Learn Linear Models In-Context.

Ruiqi Zhang, University of California, Berkeley, Berkeley, China, People's Republic of

Attention-based neural networks, particularly transformers, excel at in-context learning (ICL), where they make accurate predictions for new tasks without updating parameters given some examples. By embedding labeled training data and unlabeled test data within a prompt, transformers can perform like supervised learning algorithms. Recent studies show that transformers trained on random linear regression problems can mimic ordinary least squares (OLS) predictions. We explore ICL in transformers with a single linear self-attention layer trained via gradient flow on linear regression tasks. Despite the non-convex nature of the problem, gradient flow with appropriate initialization reaches a global minimum. At this point, the transformer achieves prediction errors comparable to the best linear predictor for new tasks. However, the study also highlights that while the trained transformer can handle certain distribution shifts, it struggles with shifts in the covariate distribution of the prompts. In a generalized ICL setting where covariate distributions vary across prompts, the transformer remains brittle under mild covariate shifts.

4 - Why Transformers Need Adam: A Hessian Perspective

Yushun Zhang, The Chinese University of Hong Kong, Shenzhen, shenzhen, China, People's Republic of

SGD performs worse than Adam by a significant margin on Transformers, but the reason remains unclear. In this work, we provide an explanation of SGD's failure on Transformers through the lens of Hessian: (i) Transformers are heterogeneous": the Hessian spectrum across parameter blocks vary dramatically, a phenomenon we callblock heterogeneity"; (ii) Heterogeneity hampers SGD: SGD performs badly on problems with block heterogeneity. To validate that heterogeneity hampers SGD, we check various Transformers, CNNs, MLPs, and quadratic problems, and find that SGD works well on problems without block heterogeneity but performs badly when the heterogeneity exists. Our initial theoretical analysis indicates that SGD fails because it applies one single learning rate for all blocks, which cannot handle the heterogeneity among blocks. The failure could be rescued if we could assign different learning rates across blocks, as designed in Adam.

5 - Accelerating Convergence of Score-Based Diffusion Models, Provably

Yu Huang, Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, PHILADELPHIA, PA, United States

Score-based diffusion models, while achieving remarkable empirical performance, often suffer from low sampling speed, due to extensive function evaluations needed during the sampling phase. Despite a flurry of recent activities towards speeding up diffusion generative modeling in practice, theoretical underpinnings for acceleration techniques remain severely limited. In this paper, we design novel training-free algorithms to accelerate popular deterministic (i.e., DDIM) and stochastic (i.e., DDPM) samplers. Our accelerated deterministic sampler converges at a rate O(1/T^2) with T the number of steps, improving upon the O(1/T) rate for the DDIM sampler; and our accelerated stochastic sampler converges at a rate O(1/T), outperforming the rate O(1/\sqrt{T}) for the DDPM sampler. The design of our algorithms leverages insights from higher-order approximation, and shares similar intuitions as popular high-order ODE solvers like the DPM-Solver-2. Our theory accommodates ℓ2-accurate score estimates, and does not require log-concavity or smoothness on the target distribution.

TD32

Summit - 423

Distributional Robustness

Invited Session

OPT: Optimization Under Uncertainty

Chair: Sloan Nietert, Cornell University, Ithaca, NY, United States

Co-Chair: Soroosh Shafiee, Cornell, Ithaca, NY, United States

1 - From Distributional Robustness to Robust Statistics

Mohammed Amine Bennouna, Massachusetts Institute of Technology, Cambridge, MA, United States

Distributionally Robust Optimization (DRO) has been commonly considered in stochastic optimization settings under no parametric assumption or prior on the out-of-sample distribution. This lack of prior typically results in very conservative estimators. We consider here the problem of statistical estimation (and stochastic optimization) when the true distribution is known to belong to some parametric family. We study how DRO approaches extend to this setting and develop new DRO estimators that provide "optimal" statistical guarantees in a certain sense. We show then that classical robust statistics approaches naturally spur from these new DRO estimators.

2 - On Tractability, Complexity, and Mixed-Integer Convex Programming Representability of Distributionally Favorable Optimization

Nan Jiang, Georgia Tech, Atlanta, GA, United States, Weijun Xie

Distributionally Favorable Optimization (DFO) is an important framework for decision-making under uncertainty, with applications across fields such as reinforcement learning, online learning, robust statistics, chance-constrained programming, and two-stage stochastic optimization without relatively complete recourse. In contrast to the traditional Distributionally Robust Optimization (DRO) paradigm, DFO presents a unique challenge-- the application of the inner infimum operator often fails to retain the convexity. In light of this challenge, we study the tractability and complexity of DFO. We establish sufficient and necessary conditions for determining when DFO problems are tractable or intractable. Despite the typical nonconvex nature of DFO problems, our findings show that they are mixed-integer convex programming representable (MICP-R), thereby enabling solutions via standard optimization solvers. Finally, we numerically validate the efficacy of our MICP-R formulations.

3 - Outlier-Robust Wasserstein Dro

Sloan Nietert, Cornell University, Ithaca, NY, United States, Ziv Goldfeld, Soroosh Shafiee

Distributionally robust optimization (DRO) is an effective approach for data-driven decision-making in the presence of uncertainty. Geometric uncertainty due to sampling or localized perturbations of data points is captured by Wasserstein DRO (WDRO), which seeks to learn a model that performs uniformly well over a Wasserstein ball centered around the observed data distribution. However, WDRO fails to account for non-geometric perturbations such as adversarial outliers, which can greatly distort the Wasserstein distance measurement and impede the learned model. We address this gap by proposing a novel outlier-robust WDRO framework for decision-making under both geometric (Wasserstein) perturbations and non-geometric (total variation (TV)) contamination that allows an ε-fraction of data to be arbitrarily corrupted. We design an uncertainty set using a certain robust Wasserstein ball that accounts for both perturbation types and derive minimax optimal excess risk bounds for this procedure that explicitly capture the Wasserstein and TV risks. We prove a strong duality result that enables tractable convex reformulations and efficient computation of our outlier-robust WDRO problem. When the loss function depends only on low-dimensional features of the data, we eliminate certain dimension dependencies from the risk bounds that are unavoidable in the general setting. Finally, we present experiments validating our theory on standard regression and classification tasks.

4 - Continuous Time Distributionally Robust Bayesian Utility Maximization

Lukas-Benedikt Fiechtner, Stanford University, Stanford, CA, United States, Jose Blanchet

We consider a utility maximization stochastic control problem in continuous time. In our market, prices are driven by unobserved factors. The controller thus faces a partially observed decision making problem. We consider a full Bayesian model for which the unobserved factors can be filtered out and the control problem can be explicitly solved. Then, we consider a distributionally robust partially observed formulation involving the adapted Wasserstein distance. The formulation is non-trivial for several reasons, but one that is conceptually interesting involves asymmetric information. Precisely, the information available to the agent and the adversary is fundamentally different (with the adversary is allowed to see information about the factors which is hidden to the agent). We present a solution for the optimal investment policy under an extension of Merton's model for portfolio selection.

TD33

Summit - 424

Methods and Applications in Stochastic and Chance-Constrained Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Yiling Zhang, University of Minnesota, Minneapolis, MN, United States

1 - On the ReLU Lagrangian Cuts for Stochastic Integer Programs

Haoyun Deng, Georgia Institute of Technology, Atlanta, GA, United States, Weijun Xie

We study stochastic integer programs where both first-stage and recourse decisions can have mixed-integer values. A new family of Lagrangian cuts, termed ``ReLU Lagrangian cuts,'' is developed by reformulating the nonanticipativity constraints using ReLU functions. These cuts can be integrated into scenario decomposition algorithms. We prove that the inclusion of ReLU Lagrangian cuts is sufficient to solve the original stochastic integer programs. Without solving the Lagrangian dual problems, we derive closed-form expressions for these cuts. Furthermore, to speed up the cut-generating procedures, we introduce linear programming-based techniques to enhance the coefficients of the cuts. Numerical studies demonstrate the effectiveness of the proposed cuts compared to existing methods.

2 - Optimization with Multi-Sourced Reference Information and Unknown Trust: a Distributionally Robust Approach

Yanru Guo, University of Michigan, Ann Arbor, MI, United States, Bo Zhou, Ruiwei Jiang, Siqian Shen

Navigating the complexities of decision-making processes relies heavily on the ability to integrate information from multiple sources, especially in environments where uncertainty exists. In this study, we investigate a stochastic optimization problem where the probability distributions of uncertain parameters are estimated using data from various sources. We conduct non-parametric data fusion based on trust over different sources to construct the Wasserstein ambiguity set and formulate a distributionally robust optimization (DRO) model to minimize potential worst-case losses. We reformulate the DRO model into a computationally tractable linear program (LP). In addition, we design a dynamic trust update framework based on actual losses under trust misalignment or data misalignment and introduce probability dominance to identify information sources with dominant trust. We demonstrate the effectiveness of our approaches using instances of portfolio optimization and resource allocation for wildfire control.

3 - Integer Programming Approaches for Distributionally Robust Chance Constraints with Adjustable Risks

Yiling Zhang, University of Minnesota, Minneapolis, MN, United States

We study distributionally robust chance-constrained programs (DRCCPs) with individual chance constraints. The DRCCPs treat the risk tolerances associated with the distributionally robust chance constraints (DRCCs) as decision variables to trade off between the system cost and risk of violations by penalizing the risk tolerances in the objective function. We consider individual chance constraints with uncertainty either on the right-hand side or on the left-hand side. By exploring the discrete structures, we develop mixed integer programming reformulations for the two types of uncertainty to determine the optimal risk tolerance for the chance constraint. Valid inequalities are derived to strengthen the formulations. We test diverse instances of diverse sizes.

TD34

Summit - 425

Evolution and Impact in Cyber-Social Environments

Contributed Session

Chair: Yasmin Merali, University of Hull, Hull, HU17 8DE, United Kingdom

1 - Deploying the Complex Adaptive Systems Paradigm to Address Cyber-Social Dynamics

Yasmin Merali, University of Hull, Hull, United Kingdom

This paper focuses on the challenges posed by the emergence of *cyber-social* phenomena across networks straddling the embodied world and cyberspace. The internet and emergent technologies resulted in a step change in the level of complexity inherent in the effective world, i.e. the world as perceived, experienced, and understood by players, constituting the context within which decisions are made, actions taken, and consequences realised. The subsequent incorporation of AI-based functionality has further accelerated generation of discontinuities in the global socio-economic and geo-political context.

These developments have resulted in an increasing dependence on *representations* of reality and a deceasing capacity for experiential or embodied participation in interactions with the effective world, precipitating a shift from:

- *social* networks to *cyber-social* networks;
- the *social construction* of the world to the *informational construction* of *multiple* worlds;
- discontinuity in the world over a *time line* to discontinuity of *contemporaneous parallel worlds.*

The pace of technological advances outstrips the ability of legislators to develop adequate safeguarding measures against mis-information and disinformation: consequently organisations and institutions must develop strategies for resilience and co-evolution in an informational landscape characterised by dynamism and uncertainty.

This paper deploys the information lens and the *Complex Adaptive Systems* paradigm to:

- define key features of the landscape, highlighting potential dangers inherent in the way that BigData and AI approaches impact on the cognitive, social and contextual filtering of information, and

- develop the conceptual scaffolding for countering these dangers, attending to the evolution of semantic and semiotic dimensions over time.

2 - Platform-Driven Collaboration Patterns: Structural Evolution Over Time and Scale

Babak Heydari, Northeastern University, BOSTON, MA, United States, Negin Maddah

This study explores the dynamics of decentralized collaboration within a digital organizational landscape, contrasting these with traditional models. We introduce a novel methodology that captures high-level collaboration dynamics, extending beyond direct messages to include both temporal and content dimensions of user interactions. This approach utilizes an Alternating Timed Interaction (ATI) metric for temporal analysis and a quantitative strategy for assessing thematic similarity via Natural Language Processing (NLP). The findings validate three hypotheses that collectively underscore the complexities of digital team dynamics within sociotechnical systems: Firstly, it establishes the significant influence of problem context on team structures in work environments, emphasizing the need to consider the specific nature of tasks in analyzing collaborative dynamics. Second, we identify significant evolving patterns of team structures (such as decentralized leadership, information accessibility, diffusion speed of ideas, and the formation of sub-communities and polarization) versus team size and artifact maturity (age). Lastly, we demonstrate notable differences in team structure patterns between digital platforms (exemplified by the editorial collaboration networks of Wikipedia articles) and traditional organizational settings (using the results of a previous study of 65 US firms). By mapping out network structures and collaborative patterns on Wikipedia, this study not only illustrates unexplored aspects of digital platform dynamics but also sets the stage for strategic interventions to optimize digital team dynamics to align with broader organizational goals. These findings are instrumental for organizations navigating the digital era.

3 - The Impact of Misinformation on Government Policy Performance: Moderating Effects Through Public Risk Perception

Meichen Lu, University of Essex, SOUTHEND-ON-SEA, United Kingdom, Maged Ali, Niraj Kumar, Wen Zhang

Influenced by subjective public risk perceptions, misinformation can either be magnified or minimized during transmission, thereby impacting the efficacy of government policies in managing crises. However, various risks perceived during crises remain relatively unexplored. Drawing on the Protection Motivation Theory, which considers diverse individual decision-making processes when confronted with risks, we proposed a conceptual model to uncover the nuanced role of risk perception during a crisis. Using the COVID-19 pandemic as the case of this paper, we tested hypotheses using data collected on perceived risks at three different levels: overall (macro-level), interpersonal (meso-level), and individual (micro-level), covering the period from February 2020 to January 2022. Consequently, we found a detrimental impact of misinformation diffusion on the effectiveness of government policies during the crisis. This deleterious impact could be further alleviated through heightened macro- and meso-level risk perceptions but exacerbated by an escalation in micro-level risk perception. The findings provide insightful discussion to elucidate the risk perception paradox: why individuals can be vulnerable to misinformation's impact on policy performance under specific circumstances yet exhibit resilience against its influence in other situations. Accordingly, we advocate for policymakers and public health authorities to craft refined and targeted risk communication strategies by considering the heterogeneous nature of the public's risk perception.

4 - How elasticity of organizational identity shape technological innovation: From Chinese special new enterprises

Yang Rong, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Youmin Xi, Jing Ge

Abstract:

In order to reveal the relationship between organizational core attributes and technological innovation, elasticity of organizational identity plays an essential role. This case study show that the orientation of technological innovation was set for balancing the renewal or conflict of value labels, through timing contrast or time point comparison of socially constructed identity. The inner tension from professional identity and innovation identity drives technological innovation models and maintains organizational identity resilience. The evolution of Professional label shapes innovative search for opportunities and product serialization strategy. As for innovation capability label, the elasticity lies in creative resources of balance, focus, coordination and leverage. Those were shaped by the harmonious coupling of "psychological capital storage and hardware support system completion", as wellas the mechanism of action manifested as the release of innovative vitality. Innovation capability elasticity reduces the volatility of innovation willingness, update technology track. Based on the technical feature and innovation environment openness, the nature of organization influence the decision on binary equilibrium of exploratory and exploitative innovation, independent and cooperative innovation. Effective mode of technological innovation further balance the dialectic tensions between innovation efficiency and innovation quality, cognitive legitimacy and uniqueness, flexibility and durability when the organizational identity is pulling apart.

5 - Design Structure Matrices and Modularity in Services

Cheryl Druehl, George Mason University, Fairfax, VA, United States

Using higher education as the context, we examine mirroring in service organizations to explore organization structure, service offerings, and internal service design processes. The mirroring hypothesis suggests modular organizations will have modular products and vice versa. Design structure matrices are used to identify modular organizations and modular offerings. The difference in methods used to design modular services in mirrored organizations versus cross functional services or integrated services is highlighted.

TD35

Summit - 427

Advancing Innovations in Mobility: International Collaboration and Technological Innovations in Transportation

Invited Session

TSL: Intelligent Transportation Systems

Chair: Yudai Honma, The University of Tokyo, Meguro-ku, 153-8505, Japan

1 - Required Battery Capacity of EVs Assuming Optimal Locations of In-motion Wireless Power Transfer System

Yudai Honma, The University of Tokyo, Meguro-ku, Japan

In-motion Wireless Power Transfer System (WPTS) is gaining attention as infrastructure to support electric vehicle (EV) mobility. This study assumes optimal locations of WPTSs in urban areas and analyzes battery consumption patterns in EV mobility. By simulating EV charging and discharging patterns while considering acceleration, deceleration, and traffic signals, the study discusses the required battery capacity. This research highlights two major novelties. First, it is demonstrated that the feasibility of EV mobility demand largely depends on the location of WPTS. Specifically, when optimally located, the total installation distance is reduced, while a uniform location at each intersection significantly increases the total installation distance required to meet mobility demands. Second, regarding battery capacity, it is shown that approximately 5kWh is sufficient when assuming optimal placement, which is significantly less than the current standard capacity. This underscores the promising potential of WPTS.

2 - From Disruption to Recovery: Enhancing Resilience for Transportation Networks Capacity

MOHAMMAD ABBASI, Arizona State University, Tempe, AZ, United States, Xuesong (Simon) Zhou

Resilience in transportation systems relies on both absorptive and adaptive capacities to mitigate disruptions and restore normal operations efficiently. A disruptive incident triggers a decline in a system capacity, which can be mitigated through strategic decision-making, such as routing and rerouting strategies. By controlling these parameters, the system can gradually recover lost capacity, depending on the severity of the disruption and the system's healing capability.

However, challenges such as accurately defining resilience metrics, incorporating emerging technologies like connected and autonomous vehicles, and balancing different objectives, such as mobility, accessibility, and environmental impact, requires integrated modeling approaches.

To address these challenges, this study explores the intersection of adaptive capacity using optimization techniques such as column generation. Integrating optimization models, including discrete choice and system dynamics models, offers a comprehensive approach to simulating disruption scenarios and evaluating system responses. Moreover, quantifying absorptive and restorative capacities enables the identification of strategies to enhance resilience. Practical applications range from informing infrastructure investments to guiding policy-making decisions.

3 - Designing Urban Road Hierarchies for Pedestrian Safety with a Mathematical Optimization Approach

Hiroyuki Hasada, The University of Tokyo, Meguro, Japan, Fuma Ikeya, Azusa Toriumi, Yudai Honma, Takashi Oguchi

Crossing or walking along high-speed roads poses a significant risk of traffic accidents for pedestrians. It is essential to improve pedestrian accessibility by properly defining road hierarchies to ensure a safe walking environment. However, especially in Japan, zone 30 areas are often limited and arbitrarily designated, which does not effectively protect pedestrians. This study aims to identify an optimal road hierarchy that facilitates pedestrian-friendly urban environments through mathematical theory. We formulate an optimal location problem that enables pedestrians to safely reach their destinations while maintaining vehicle travel to a certain degree of smoothness. Our proposed model is the first to position this optimal location problem within the context of the inverse shortest path problem, which simultaneously finds the shortest paths for given origin-destination pairs and assigns discrete link costs. The assigned road hierarchy can inform the creation of traffic safety zones and the strategic placement of physical devices such as speed humps and chicanes. Furthermore, it is expected to be useful in evaluating how new services, such as micromobility and autonomous driving services, can expand safe activity areas for residents. Our study provides a comprehensive framework for enhancing pedestrian safety and accessibility in urban areas, ultimately contributing to the creation of more livable, pedestrian-friendly and sustainable cities.

4 - Optimizing Real-Time Transportation Operations of Reusable Packages: A Dynamic Vehicle Routing Approach Considering Managerial and Operational Factors

Monireh Mahmoudi, Michigan State University, East Lansing, MI, United States, Irandokht Parviziomran, Alireza Boloori

Transporting reusable packages involves delivering full packages from depots to customers' doors and picking up empty ones, usually modeled as a vehicle routing problem with backhauls (VRPBs). Traditional static VRPBs assume all input data, like customer requests and locations, are known beforehand. However, the rise of on-demand services requires real-time operations where data is updated during transportation, leading to dynamic VRPBs (DVRPBs). Our study utilizes an analytical framework for DVRPBs, which dynamically incorporates newly revealed customer information, decides on request acceptance/rejection based on constraints, and updates pre-planned routes accordingly. Through numerical experiments, we analyze the trade-off between transportation cost and responsiveness under various factors like urbanism, customer accessibility, dynamism degree, vehicle capacity, request time frame, and decision epoch length.

5 - Strategic Optimization of Logistics Networks for Next-Generation Mobility in Japan

Daisuke Watanabe, Tokyo University of Marine Science and Technology, Koto, Tokyo, Japan

This research explores the reform of Japan's logistics by strategically optimizing logistics networks to incorporate emerging technologies such as truck platooning and autonomous trucks. The study formulates the optimal location of logistics hubs for line haul transport in Japan using the hub location problem, considering cost reductions from increased capacity and automation due to truck platooning implementation between hubs. The findings indicate that, under various automation scenarios, as the number of hubs increases, they tend to be located in areas with high traffic demand along the Pacific Ocean. Furthermore, despite technological advancements, the study found no significant differences in location results, suggesting that considering base allocation as a single allocation is approximately feasible.

6 - Decarbonizing freight through automation and connectivity: challenges and opportunities

Prasad Gupte, Dept. of Transportation, washington, DC, United States

Although vehicles for freight are a small fraction of total vehicles in the transportation system, they constitute a disproportionately large impact of energy use and emissions, and this share has been increasing. This talk will present the technical challenges and opportunities that advances in connectivity and automation can have to decarbonize freight while also improving overall system resilience through enhanced network infrastructure and integration.

TD36

Summit - 428

Behavioral-informed Analytics and Decision-Making for Urban Transportation Systems 2

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Tianming Liu, University of Michigan, Ann Arbor, United States

Co-Chair: Manzi Li, University of Michigan, Ann Arbor, N/A, China, People's Republic of

1 - Economic Analysis of Order Consolidation and Congestion in E-commerce Delivery Service

Manzi Li, University of Michigan, Ann Arbor, China, People's Republic of, Yafeng Yin, Xin Li, Zhuoye Zhang

With the rise of e-commerce, delivery vehicles have increasingly contributed to urban traffic congestion. Despite this, limited attention has been given to the role of e-commerce deliveries in exacerbating congestion, complicating efforts to implement effective congestion pricing and reduction strategies. This study addresses this gap by developing a two-sided market framework for e-commerce systems from a queuing theory perspective, focusing on order consolidation—a key factor that balances customer demand, driver supply, road congestion, and system performance.

Theoretical analysis shows that platforms and social planners align on an optimal consolidation size that maximizes system throughput given delivery fee and driver compensation, reducing the complexity of the problem to a coupled optimization. With this property, we reveal that regulatory interventions are needed to adjust consolidation sizes when congestion is accounted for. Furthermore, the gap between optimal delivery fees and driver compensation widens in the congested scenario, potentially affecting the platform's profitability. This research offers policy insights into optimizing delivery fees, driver compensation, and consolidation strategies to manage congestion in e-commerce logistics.

2 - Integrating An Agent-Based Behavioral Model In Revenue Management for Microtransit: a Case Study In Arlington, Tx

Xiyuan Ren, New York University, Brooklyn, NY, United States, Joseph Chow

Microtransit offers a flexible and efficient solution to promote affordable and sustainable mobility services. Arlington Via provides microtransit services across the city with weekly and monthly Ride Pass plans. The service has been put into operation and generated valuable data for analysis. This study utilizes Arlington Via’s backend data and Replica’s synthetic population data to identify user dynamics. A behavioral model framework is developed wherein (1) real-time microtransit performance is derived using a calibrated simulation tool based on observed travel demand; (2) an agent-based mixed logit (AMXL) model is applied to estimate mode choice parameters at individual, origin-destination (OD) pair, and generic levels, capturing taste heterogeneity and ensuring robust predictions; and (3) the utility functions from the lower-level mode choice model are seamlessly integrated into an upper-level model to predict Ride Pass subscription. The proposed model framework is applied to evaluate operating strategies regarding operating hour, number of vehicles, discounts on specific hours, and various Ride Pass plans. Microtransit ridership, total revenue, decreased vehicle-miles-traveled (VMT), and other metrics are analyzed to compare different scenarios, providing insights to support microtransit service design.

3 - What Mobile Phone Data Reveal About Mobility Patterns of Teleworkers

Tianxing Dai, Northwestern University, Evanston, IL, United States, Peeter Kivestu, Ying Chen, Marco Nie

In a short period, the COVID-19 pandemic has transformed telework into a common practice for a significant portion of the workforce. This shift has profound implications for land use, urban development, and transportation. Traditional survey-based methods for tracking these changes are struggling to keep pace with the rapidity of this transformation. Here, we propose a method to identify different types of workers from mobile phone data, which allows us to closely examine the correlation between work arrangements, mobility patterns and key socio-demographic attributes. By applying a hierarchical clustering algorithm to a set of features extracted from a mobile phone data set, six different work types are identified and their validity is confirmed using different approaches. We find teleworkers tend to travel slower than regular workers but faster than non-workers. They also travel a shorter distance to reach their primary activity location than regular workers, but a longer distance to reach other activity location than both regular and non-workers. Our regression analysis further shows that, largely in agreement with findings in literature, racial minority and low income groups are less likely to telework.

4 - Position Lottery as An Alternative to Dynamic Pricing for Terminal Ride-Sourcing Driver Queue Control

Tianming Liu, University of Michigan, Ann Arbor, MI, United States, Yafeng Yin, Vijay Subramanian

Urban transportation terminals are hot spots for ride-hailing services and often suffer from significant supply-demand imbalance. This research aims to alleviate this issue using an innovative queue discipline control scheme on the idle driver inflow. We designed a entry position queue control as an alternative to the controversial dynamic pricing, and unveil novel insights into the optimal control properties of such strategy and design a mathematical programming framework that enables efficient computation of optimal policy in ride-sourcing operations. Numerical examples shows that the the proposed position lottery can achieve similar system performance as optimal dynamic pricing.

TD37

Summit - 429

Optimization and Learning in Recommender Systems

Contributed Session

Chair: Ziye Wu, N/A

1 - Robust portfolio optimization for recommender systems considering uncertainty of estimated statistics

Tomoya Yanagi, University of Tsukuba, Tsukuba-Shi, Japan, Shunnosuke Ikeda, Yuichi Takano

Recommender systems aim to suggest unknown but desirable items to users by predicting their preferences based on their past ratings for other items. Although previous research has often focused on improving prediction accuracy, it is also necessary to emphasize diversity and novelty in recommendations. One approach to balancing the accuracy and diversity of recommendations is the application of portfolio optimization models originally used for financial asset selection. These models consider both risk and return and have been shown to have the potential to improve both the accuracy and diversity of recommendations. However, previous studies using portfolio optimization models in recommendation systems take no account of the uncertainty of estimated statistics. When applying these models, we need to estimate user ratings and their covariances; however, these statistics are subject to unavoidable estimation errors. We thus propose a robust portfolio optimization model to determine an optimal set of recommended items by considering potential errors in the estimated statistics. Numerical experiments using public rating datasets demonstrate that our method improves not only the prediction accuracy but also the diversity of recommendations.

2 - Learning Decisions Offline from Censored Observations with ε-insensitive Operational Costs

Teng Huang, Sun Yat-sen University, Guangzhou, China, People's Republic of, Minxia Chen, Ke Fu, Miao Bai

Many important managerial decisions are made based on censored observations. Making decisions without adequately handling the censoring leads to inferior outcomes. We investigate the data-driven decision-making problem with an offline dataset containing the feature data and the censored historical data of the variable of interest without the censoring indicators. Without assuming the underlying distribution, we design and leverage ε-insensitive operational costs to deal with the unobserved censoring in an offline data-driven fashion. We demonstrate the customization of the ε-insensitive operational costs for a newsvendor problem and use such costs to train two representative machine learning models.

3 - Dynamic Asset Pricing and Allocation with Expert Forecasts

Anas Abdelhakmi, National University of Singapore, Singapore, Singapore, Andrew Lim

In this work, we explore a Bayesian inventory control framework to analyze the impact of hedging opportunities on decision-making policies when demand distribution is unknown and learned over multiple periods. We examine a multi-period Newsvendor problem with lost sales, where observed demand is right censored by inventory levels. The decision-maker faces a trade-off: overstocking to better learn the demand at the risk of surplus, or maintaining single-period optimal inventory levels at the expense of learning opportunities.

We introduce hedging opportunities, which gives the decision-maker the option to purchase the product at an additional cost, thereby reducing the cost incurred in case of overstocking. We analyze how the presence of these opportunities influence both the optimal policy and learning dynamics, showing that despite an initial cost increase, the decision-maker opts for these opportunities as they facilitate ordering additional inventory, therefore accelerating the learning process and enhancing future performance. We derive an explicit characterization for the optimal policy when demand is exponential and the rate is unknown and learned over time. Overall, the paper shows that hedging leads to quicker learning and reduced cost compared to classical models.

4 - Better and Faster Decisions with Recommendation Algorithms

Ziye Wu, National University of Singapore, Singapore, Singapore, Yiting Chen, Songfa Zhong

While recommendation algorithms have been increasingly used in daily life, little has been done to investigate their effect on decision making in terms of decision quality and preferences. Here we examine this question in an experimental setting whereby subjects from a representative US sample are randomly assigned to five conditions and make sets of binary choices between two lotteries. The two control conditions provide either no recommendations or recommendations based on a randomization device. The three treatment conditions provide recommendations developed by algorithms: one relies on choices made by the majority, while the other two employ AI-based recommenders —one based on subjects' past decisions and the other on decisions made by similar subjects. We find that subjects tend to follow recommended choices and are willing to pay a small fee to receive recommendations for their subsequent decisions. Compared to the control conditions, recommendations help to reduce noise in decisions, align preferences with the expected utility model, and accelerate decisions. These results can be explained by some classes of stochastic choice models. Our work adds to the growing literature on the behavioral underpinnings of algorithms including AI and sheds light on the design of choice architecture for decision making under risk.

TD38

Summit - 430

Air Navigation Service Provision and Capacity Planning

Contributed Session

Aviation Applications

Chair: Kam NG, The Hong Kong Polytechnic University

1 - Contrasting Risk-Averse Investment Decisions: A Comparative Study of Real Options Analysis versus Risk Measures Optimization for Airport Facilities Capacity Expansion

Ziyue Li, Florida State University, Tallahassee, FL, United States

In the realm of investment decisions, the optimal strategies are heavily influenced by the uncertain nature of demand growth. This uncertainty renders demand growth a stochastic process, leading decision-makers to adopt risk-averse approaches in their decision-making processes. There are two primary methodologies in investment decision making under uncertainty: optimization, utilizing various risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CvaR); and real options analysis, where the actual probabilities of payoffs are replaced by risk-neutral probabilities. Additionally, Knightian uncertainty can also be integrated into real options analysis to capture decision-makers' pessimistic and optimistic beliefs.

Despite the prominence of these methodologies, there remains a notable gap in the literature concerning a comparative analysis of these two approaches in investment decision-making. This study aims to fill this gap by comparing their results through an example in the context of airport facilities capacity expansion. Specifically, it explores the relationship between these approaches in determining optimal investment timing and expansion size for airport facilities capacity expansion problems, where demand follows a geometric Brownian motion (GBM) and adheres to a log-normal distribution for any given time *t*. The objective is to maximize cumulative cost savings through airport facilities capacity expansion, and the decision variables are investment timing and expansion size. Through a comprehensive analysis, this study reveals consistent results between real options analysis and risk measures optimization, shedding light on their efficacy and implications for investment decision-making in uncertain environments.

2 - Modeling Competition Between Service Providers in Urban Air Mobility

Victor Qin, MIT, Cambridge, MA, United States

Current proposals for advance air mobility (AAM) propose that third-party service providers (SPs) will be responsible for providing traffic management services for these aircraft. This paper examines two aspects of these proposals, namely that: 1) multiple SPs can operate in the same region of airspace, competing for the patronage of AAM fleet operators, and 2) SPs can themselves be fleet operators. To this end, we propose and analyze a three-tier economic model of competition between traffic management SPs. We consider several potential scenarios, including a monopolistic SP, multiple competing SPs, the presence of a public option, and a service provider who is also a fleet operator. Our analysis suggests that in the absence of excessive congestion, an SP can also be a fleet operator without significantly distorting the economic outcomes. Furthermore, we find that under certain reasonable assumptions, the presence of a public option SP can improve consumer surplus while at the same time allowing private SPs to be profitable.

3 - Terminal airspace trajectory generation with design-transferable generative models

Xinting Zhu, City University of Hong Kong, Kowloon, Hong Kong, Yongchao Ye, Lishuai Li

Generating realistic trajectories is essential for optimizing fuel economy, minimizing environmental affects, and enhancing predictive planning of airline operations, particularly within terminal airspace. This space is typically navigated via standard routes, yet actual trajectories can diverge due to factors such as air traffic control directives, pilot decisions, and aircraft performance variations. Traditional simulation methods for generating these trajectories are labor-intensive and require extensive setup for new scenarios, while existing data-driven approaches fail to adapt to new airspace designs or terminal areas lacking historical trajectory data. This study introduces a digital framework employing design-transferable generative models for terminal arrival trajectory generation. Two typical generative models with design-conditional generation, diffusion model and variational autoencoder, are employed to model the uncertainty and variability in aircraft trajectories, with integrating design profile learning. Experiments conducted at Hong Kong International Airport (HKIA) demonstrate the superior performance of our models. The results confirm the effectiveness of generative models in capturing the physical constraints and dynamics of aircraft, offering a promising, adaptable, and environmentally considerate solution for optimizing terminal airspace design and operations.

4 - Runway capacity improvement and joint TMA optimisation under data-driven dynamic wake separation prediction

Kam NG, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

Runway capacity is one leading bottleneck in near-ground flight operational efficiency, as a result of a variety of operational constraints, including aircraft wake separation and runway occupancy time. “Improving capacity and efficiency” has emerged as a critical concern with the continual surge in air travel demand and the restricted potential to expand runway configuration and other infrastructure. Compared to distance-based aircraft wake separation standard regulated by International Civil Aviation Organization (ICAO), both European Union Aviation Safety Agency (EASA) and Federal Aviation Administration (FAA) proposed the aircraft type recategorization to relax wake separation distance. Furthermore, dynamic separation that aims in weather- and aircraft-related separation, is an ongoing hot issue.
Based on our previous studies of aircraft wake vortex decay prediction, two levels of crosswind-related wake separation matrixes are proposed in this paper. We further build the runway arrival sequencing and scheduling model to explore the potential of runway capacity improvement under proposed dynamic wake separation matrixes. Both theoretical and operational runway capacity at the Hong Kong International Airport are investigated using historical arrival flight data. At the meantime, we jointly verify the possibility of arrival flight scheduling in Terminal Manoeuvre Area for supporting this separation-reduced efficient runway management. The results indicate that compared to the traditional ICAO and RECAT-EU standard, the hourly runway capacity can be improved by approximately 10% with implementing the dynamic wake separation standards through integrated TMA and runway optimisation.

TD39

Summit - 431

AAS Distinguished Speaker

Invited Session

Aviation Applications

Chair: Nuno Ribeiro, Singapore University of Technology and Design, Singapore, Singapore

Co-Chair: Lishuai Li, City University of Hong Kong, Kowloon, Hong Kong

1 - The Future of Last-Mile Delivery Technology and the Importance of a Holistic Approach

Yuqian Dong, Walmart, Seattle, WA, United States

As the global e-commerce market grows, the demand for efficient last-mile delivery systems has intensified, making investment in emerging technologies essential for staying competitive. Key challenges include high costs, logistical complexity, environmental impact, and rising customer expectations. Investing in technologies such as drones, autonomous delivery vehicles, micro-fulfillment centers, and AI-powered tracking systems is critical for reducing delivery times, lowering costs, and improving sustainability. Successful adoption of these technologies must be approached holistically. It is crucial to integrate these innovations across the entire last-mile delivery ecosystem, ensuring that technological advancements align with broader business goals, optimize operations, reduce costs, and meet consumer demands for speed and transparency. Moreover, as we adopt these technologies, it is vital to mitigate potential adverse impacts on safety, legal compliance, and society at large. Ensuring that these innovations are introduced in a way that prioritizes safety, adheres to legal standards, and considers societal implications is crucial for sustainable growth and long-term success. Real-world examples from industry leaders like Amazon and Walmart demonstrate how early, responsible investments in both technology and holistic integration are already transforming last-mile delivery. A future where these responsible, integrated solutions drive a smarter, more sustainable, and customer-centric approach to logistics positions academics and businesses to not only adapt to changes but also to actively shape the future of delivery.

TD40

Summit - 432

Emerging Topics in Behavioral Operations Management

Invited Session

Behavioral Operations Management

Chair: Behrooz Pourghannad, Lundquist College of Business, University of Oregon, Eugene, OR, United States

Co-Chair: Yuanyuan Ding, University of Oregon, Eugene, OR

1 - Behavioral Response to Priority in Deceased-Donor Kidney Allocation

Jiayi Liu, Virginia Tech, BLACKSBURG, VA, United States, Diwas KC

This study examines the impacts of priority in deceased-donor kidney allocation on patient outcomes. Using national kidney transplant data and a regression-discontinuity design, we show that allocation priority generates a positive supply shock. This supply shock not only enhances health outcomes for transplant recipients but also leads to unintended behavioral side effects.

2 - The Hidden Cost of Hidden Fees: a Dynamic Analysis of Price Obfuscation in Online Platforms

Jose Lopez, MIT, Cambridge, MA, United States, Edward Anderson

We study the effects of a common price obfuscation tactic, “shrouding hidden fees,” on consumer behavior and platform firm performance. Where traditional economic models of individual firms have shown that obfuscation tactics can be profitable for these firms even in repeated interactions, more recent work in behavioral operations management has argued that these tactics can be harmful not just to consumers but to the firms themselves. We contribute to these studies by explicitly accounting for different aspects of platform value creation to understand the role and incentives of platform firms as intermediaries that facilitate the matching process, and by using simulation modeling methods to expand model boundaries, and study appropriate time horizons. We find evidence to suggest that building consumer trust through disclosure is a dynamic attribute characterized by worse-before better outcomes. The results provide evidence that platform pricing transparency decisions may evolve differently depending on market and industrial context.

3 - How the consumers respond to responsibility initiatives in the developing world?

Neslihan Ozlu, Georgia Tech, Scheller College of Business, Atlanta, GA, GA, United States, Basak Kalkanci

With an increasing focus on sustainability, firms are starting to assess how their social responsibility efforts influence consumers. While studies evaluating consumer reaction to corporate social responsibility predominantly focus on US consumers, many firms are expanding their presence and seeking to engage effectively with consumers in developing economies. Examining the relevance of social responsibility for consumers in the developing world is particularly crucial since significant negative social impacts occur in upstream supply chains typically located in these countries (which may increase consumer sensitivity to responsibility), while consumers in these countries are also highly prone to economic concerns (potentially reducing their sensitivity to responsibility). Thus motivated, we conducted field experiments in collaboration with a major apparel retailer in Turkey to analyze the link between social responsibility initiatives and consumer purchase behavior. Our experiments leverage the digital infrastructure of retail stores to provide transparency into the firm's social responsibility initiatives through audio and videos at different times. We highlight internal (e.g., engaging women in economically disadvantaged areas as direct supply chain partners to produce handcrafted products) and external (helping women in those areas learn handicraft skills) responsibility initiatives and use transaction data to link experimental conditions to sales. We find that transparency into responsibility initiatives tends to positively affect sales, particularly when coupled with a discount. Furthermore, local socio-economic profile and store characteristics tend to moderate this link. This study underlines the importance of creating locally tailored strategies for prioritizing and communicating social responsibility activities for firms.

TD41

Summit - 433

Network Data Insights: Algorithmic Approaches

Invited Session

Applied Probability Society

Chair: Yeganeh Alimohammadi, UC Berkeley, Berkeley, CA, United States

1 - A Method for Experiments Under Unknown Network Interference

Sadegh Shirani, Stanford University, Stanford, CA, United States, Mohsen Bayati

Randomized experiments are a powerful methodology for data-driven evaluation of decisions or interventions. Yet, their validity may be undermined by network interference. This occurs when the treatment of one unit impacts not only its outcome but also that of connected units, biasing traditional treatment effect estimations. Our study introduces a new framework to accommodate complex and unknown network interference, moving beyond specialized models in the existing literature. We present a practical algorithm for estimating the treatment effect and demonstrate its efficacy in multiple numerical scenarios.

2 - Quantitative Propagation of Chaos for Non-Exchangeable Diffusions

Lane Chun Yeung, Carnegie Mellon University, Pittsburgh, PA, United States, Daniel Lacker, Fuzhong Zhou

This talk discusses an ongoing work on quantitative propagation of chaos for non-exchangeable systems. For systems of diffusions with heterogeneous interactions, we obtain a variety of entropy bounds quantifying the degree of approximate independence of the particles in terms of the fine structure of the matrix of pairwise interaction strengths. Our proof strategy is based on adapting a recent approach for exchangeable systems, which relies on a hierarchy of differential inequalities bounding the k-particle entropy in terms of the $k+1$-particle entropy. The hierarchy is significantly more complex in the non-exchangeable setting, indexed by sets rather than numbers of particles, and we analyze it by means of an unexpected connection with first-passage percolation. Joint work with Daniel Lacker and Fuzhong Zhou.

3 - From Signaling to Interviews in Random Matching Markets

Sophie Yu, University of Pennsylvania, Philadelphia, PA, United States, Itai Ashlagi, Amin Saberi, Maxwell Allman

In many two-sided matching markets, agents initially engage in costly interviews in order to refine their interim preferences. We study how signaling mechanisms can reduce the number of interviews in random markets. We are interested in perfect interim stability, which expands the notion of stability to ensure that no two agents regret not interviewing each other. The final match is almost interim stable if it is perfect interim stable after removing a vanishingly small fraction of agents.

For single-tiered markets, we study the impact of short-side (resp. long-side, both-side) signaling, where agents from short-side (resp. long-side, both-side) of the market signal their top d preferred candidates. The interview graphs are formed by including all pairs where at least one party has signaled to the other. When $d=\omega(1)$, we show that short-side signaling leads to almost interim stable matchings for every stable matching. Long-side signaling is effective in weakly imbalanced markets (where the difference in size between market sides is small), but fails in strongly imbalanced markets (where this difference is significant). We also demonstrate the failure of both-side signaling when the impact of interviews is negligible with $d=o(\log n)$. When $d\ge \Omega(\log^2 n)$, short-side signaling achieves perfect interim stability, while long-side signaling fails in imbalanced markets. We extend our results to multi-tiered markets, proposing a signaling mechanism that achieves interim stability across tiers. Our analysis also identifies conditions for truthful signaling for our signaling mechanisms across various market conditions.

4 - Electric Robo-Taxi Dispatching and Charging Facility Allocation Using Deep Reinforcement Learning

Zhanhao Zhang, Cornell University, Ithaca, NY, United States, Manxi Wu, Jim Dai

The adoption of autonomous driving technologies, particularly through robo-taxi services, is revolutionizing the transportation industry. Companies like Cruise and Waymo are deploying robo-taxis in several US cities, including San Francisco, Phoenix, and Austin. These robo-taxis, being electric vehicles, present operational challenges since they take longer to charge compared to traditional gasoline cars. Consequently, ride-hailing companies must optimize battery charging schedules and deploy charging infrastructures in accordance with trip demand and fleet distribution. In this work, we model the control of an electric vehicle (EV) fleet as a Markov decision process with an infinite horizon and a discrete state space. The joint control of an EV fleet is challenging because the state and action space grows exponentially with the number of EVs. To address this, we introduce a scalable control policy optimization pipeline that employs atomic action decomposition, state aggregation, and neural network function approximation. We prove that atomic action decomposition is without loss of optimality. Additionally, we propose a fluid-based linear programming (LP) approach that provides an upper bound on the long-run average daily revenue achievable by any control policy. Numerical experiments benchmarking against the objective value obtained from the fluid-based LP using the NYC For-Hire Vehicle dataset demonstrate the superb performance of our pipeline.

5 - Characterizing Tie Strength with An Algebraic Topological Stochastic Process

Arnab Sarker, Massachusetts Institute of Technology, Cambridge, MA, United States, Jean-Baptiste SEBY, Austin Benson, Ali Jadbabaie

The association between tie strength and social structure is a fundamental topic in the social sciences. This work investigates this association by analyzing tie strength in higher-order networks, an increasingly relevant model which can encode group interactions between three or more individuals. Specifically, in this talk I will compare tie strength to Edge PageRank, an algebraic topological extension of the classical PageRank measure for higher-order networks. I will first show how the stochastic process underlying Edge PageRank can be understood as a random process which models communication flow in the social network, motivating the measure's use in estimating tie strength. I will then provide a characterization of the Edge PageRank measure in terms of the Hodge decomposition, a tool from algebraic topology which can be shown to be interpretable in the context of social networks. Empirically, analysis across nine datasets reveals Edge PageRank is strongly inversely correlated with tie strength and outperforms traditional measures in predicting tie strength based on network structure. Additionally, Edge PageRank emphasizes ties that span a medium network distance and likely have informational advantages, ultimately enabling the measure to capture population-scale phenomena related to the strength of long ties. These results indicate that ties associated with higher-order interactions tend to be stronger, suggesting that group interactions reinforce strong ties -- a result not easily captured by traditional network models.

TD42

Summit - 434

Experimentation With Interference

Invited Session

Applied Probability Society

Chair: Peter Frazier, Cornell / Uber, Ithaca, NY, 14850, United States

Co-Chair: su jia, Cornell, New York, NY, United States

1 - Causal Inference under Stochastic Congestion

Stefan Wager, Stanford GSB, Stanford, CA, United States

Whenever one runs randomized experiments in a service system, stochastic congestion can arise from temporarily limited supply and/or demand. Such congestion gives rise to cross-unit interference between the waiting customers, and analytic strategies that do not account for this interference may be biased. In this talk, I will survey some recent advances on causal inference in settings with stochastic congestion.

2 - Experimentation under Spatio-temporal Interference

su jia, Cornell, New York, NY, United States

Experimentation with interference poses a significant challenge in contemporary online platforms. Prior research on experimentation with interference has concentrated on the final output of a policy. The cumulative performance, while equally crucial, is less well understood. To address this gap, we introduce the problem of {\em Multi-armed Bandits with Interference} (MABI), where the learner assigns an arm to each of N experimental units over a time horizon of T rounds. The reward of each unit in each round depends on the treatments of {\em all} units, where the influence of a unit decays in the spatial distance between units. Furthermore, we employ a general setup wherein the reward functions are chosen by an adversary and may vary arbitrarily across rounds and units. We first show that switchback policies achieve an optimal {\em expected} regret $\tilde O(\sqrt T)$ against the best fixed-arm policy. Nonetheless, the regret (as a random variable) for any switchback policy suffers a high variance, as it does not account for N. We propose a cluster randomization policy whose regret (i) is optimal in {\em expectation} and (ii) admits a high probability bound that vanishes in N.

3 - Interference in Matching Marketplaces: From Estimation to Optimization

Arthur Delarue, Stewart School of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA, United States, Kleanthis Karakolios

Marketplace companies routinely use randomized experiments to make operational decisions. Randomized controlled trials are often used to decide whether a specific intervention should be rolled out. However, marketplace experiments suffer from interference, where the treatment status of one unit can affect the outcome of another unit. In the first part of this talk, we consider interventions which stimulate demand but reduce the value of each unit to the platform (discounted demand nudges). We show that the standard estimators exhibit significant bias, which may be positive or negative depending on the experiment design. We then propose alternative, bias-reducing estimators based on linear programming shadow prices. Finally, we consider how these various estimators can influence rollout decisions of particular interventions.

4 - Price Experimentation and Interference

Orrie Page, Stanford, Stanford, CA, United States, Wassim Dhaouadi, Ramesh Johari, Gabriel Weintraub

In this paper, we examine biases arising in A/B tests where firms modify a continuous parameter, such as price, to estimate the global treatment effect of a given performance metric, such as profit. These biases emerge in canonical experimental estimators due to interference among market participants. We employ structural modeling and differential calculus to derive intuitive characterizations of these biases. We then specialize our general model to a standard revenue management pricing problem. This setting highlights a key pitfall in the use of A/B pricing experiments to guide profit maximization: notably, the canonical estimator for the expected change in profits can have the wrong sign. In other words, following the guidance of canonical estimators may lead firms to move prices in the wrong direction, inadvertently decreasing profits relative to the status quo. We apply these results to a two-sided market model and show how this “change of sign" regime depends on model parameters such as market imbalance, as well as the price markup. Finally, we discuss structural and practical implications for platform operators.

TD43

Summit - 435

Generative AI for Decision-making

Invited Session

Applied Probability Society

Chair: Guanting Chen, UNC Chapel Hill, Chapel Hill,, NC, United States

Co-Chair: Xiaocheng Li, Imperial College Business School, London, United Kingdom

1 - Reason for Future, Act for Now: A Principled Framework for Autonomous LLM Agents with Provable Sample Efficiency

Zhihan Liu, Northwestern University, Evanston, IL, United States, Zhaoran Wang

Large language models (LLMs) demonstrate impressive reasoning abilities, but translating reasoning into actions in the real world remains challenging. In particular, it remains unclear how to complete a given task provably within a minimum number of interactions with the external environment, e.g., through an internal mechanism of reasoning. To this end, we propose a principled framework with provable regret guarantees to orchestrate reasoning and acting, which we call "reason for future, act for now" (RAFA). Specifically, we design a prompt template for reasoning that learns from the memory buffer and plans a future trajectory over a long horizon ("reason for future"). At each step, the LLM agent takes the initial action of the planned trajectory ("act for now"), stores the collected feedback in the memory buffer, and reinvokes the reasoning routine to replan the future trajectory from the new state. The key idea is to cast reasoning in LLMs as learning and planning in Bayesian adaptive Markov decision processes (MDPs). Correspondingly, we prompt LLMs to form an updated posterior of the unknown environment from the memory buffer (learning) and generate an optimal trajectory for multiple future steps that maximizes a value function (planning). The learning and planning subroutines are performed in an "in-context" manner to emulate the actor-critic update for MDPs. Our theoretical analysis proves that the novel combination of long-term reasoning and short-term acting achieves a squared-root regret. In particular, the regret bound highlights an intriguing interplay between the prior knowledge obtained through pretraining and the uncertainty reduction achieved by reasoning and acting. Our empirical validation shows that it outperforms various existing frameworks and achieves nearly perfect scores on a few benchmarks. By incorporating "classical" MDP techniques, RAFA introduces the first autonomous LLM agent with provable regret guarantees.

2 - Transformers as Decision Makers: Provable In-Context Reinforcement Learning via Supervised Pretraining

Licong Lin, UC Berkeley, Berkeley, CA, United States, Yu Bai, Song Mei

Large transformer models pretrained on offline reinforcement learning datasets have demonstrated remarkable in-context reinforcement learning (ICRL) capabilities, where they can make good decisions when prompted with interaction trajectories from unseen environments. However, when and how transformers can be trained to perform ICRL have not been theoretically well-understood. In particular, it is unclear which reinforcement-learning algorithms transformers can perform in context, and how distribution mismatch in offline training data affects the learned algorithms. This paper provides a theoretical framework that analyzes supervised pretraining for ICRL. This includes two recently proposed training methods -- algorithm distillation and decision-pretrained transformers. First, assuming model realizability, we prove the supervised-pretrained transformer will imitate the conditional expectation of the expert algorithm given the observed trajectory. The generalization error will scale with model capacity and a distribution divergence factor between the expert and offline algorithms. Second, we show transformers with ReLU attention can efficiently approximate near-optimal online reinforcement learning algorithms like LinUCB and Thompson sampling for stochastic linear bandits, and UCB-VI for tabular Markov decision processes. This provides the first quantitative analysis of the ICRL capabilities of transformers pretrained from offline trajectories.

3 - Towards Understanding the Learning and Generalization of Decision Transformers

Guanting Chen, UNC Chapel Hill, Chapel Hill,, NC, United States

We consider the supervised pretrained transformer for a class of sequential decision-making problems. The class of considered problems is a subset of the general formulation of reinforcement learning in that there is no transition probability matrix, and the class of problems covers bandits, dynamic pricing, and newsvendor problems as special cases. Such a structure enables the use of optimal actions/decisions in the pretraining phase, and the usage also provides new insights for the training and generalization of the pretrained transformer.

4 - Uncertainty Estimation and Quantification for LLMs: A Simple Supervised Approach

Linyu Liu, UNC Chapel Hill, Chapel Hill, NC, United States, Yu Pan, Xiaocheng Li, Guanting Chen

Large language models (LLMs) are highly capable of many tasks but they can sometimes generate unreliable or inaccurate outputs. To tackle this issue, this paper studies the problem of uncertainty estimation and calibration for LLMs. We begin by formulating the uncertainty estimation problem for LLMs and then propose a supervised approach that takes advantage of the labeled datasets and estimates the uncertainty of the LLMs' responses. Based on the formulation, we illustrate the difference between the uncertainty estimation for LLMs and that for standard ML models and explain why the hidden activations of the LLMs contain uncertainty information. Our designed approach effectively demonstrates the benefits of utilizing hidden activations for enhanced uncertainty estimation across various tasks and shows robust transferability in out-of-distribution settings. Moreover, we distinguish the uncertainty estimation task from the uncertainty calibration task and show that a better uncertainty estimation mode leads to a better calibration performance. In practice, our method is easy to implement and is adaptable to different levels of model transparency including black box, grey box, and white box, each demonstrating strong performance based on the accessibility of the LLM's internal mechanisms.

TD44

Summit - 436

Optimizing Healthcare and Equity through Location Analytics

Contributed Session

Chair: Zhiyuan Wei, California Polytechnic State University, San Luis Obispo, Buffalo, NY, United States

1 - Locating Hope: Harm Reduction Smart Vending Machines Location Problem

Sara Abu Aridah, The Pennsylvania State University, University Park, PA, United States, Paul Griffin

Efficient allocation of resources for harm reduction initiatives in substance use disorder communities is crucial for mitigating the adverse impacts of drug addiction. This study focuses on developing a mathematical model to optimize the placement of Smart Vending Machines (SVMs), aimed at reducing harm associated with substance abuse while minimizing operational costs. The model considers a set of communities, each with a candidate location for an SVM, and incorporates community-level data on substance abuse indicators to inform decision-making. Demand within each community arises from both individuals directly affected by Opioid Use Disorder (OUD) and those within their social network or from concerned community members. Furthermore, SVM locations are subject to capacity constraints to ensure effective resource utilization.

We use a mixed integer linear mathematical model that integrates operations research techniques with insights from public health to address the complex challenges of harm reduction intervention. By strategically positioning SVMs based on community-level demand and operational constraints, the model aims to maximize harm reduction impact while optimizing resource allocation. Monte Carlo simulations are employed to account for uncertainty in demand, enhancing the robustness of the model's solutions.

This research contributes to the advancement of evidence-based approaches to harm reduction by providing decision-makers with a systematic framework for deploying SVMs in substance abuse communities. The findings hold implications for public health policy and intervention strategies, offering insights into the effective allocation of resources to combat substance abuse and its associated harms.

3 - Can Combinatorial Equity Models Solve Large Problem Instances?

Emily Speakman, University of Colorado Denver, Denver, CO, United States, Daphne Skipper, Drew Horton, Tom Logan

Equitable Facility Location is the problem of optimally locating vital services within a community so that equitable access is prioritized. The discussion of equity is vital because low-income neighborhoods and communities of color are consistently shown to be disproportionately impacted by poor access to amenities. Issues of equitable access are well-studied in the operations literature with several metrics having been proposed for optimization. However, a consensus on the ideal measure of equity has been elusive, in part, because metrics able to capture equity well result in highly nonlinear optimization models that are computationally expensive to solve. In this talk, we present recent work that offers a solution to this challenge and adds to the continued discussion of equity in facilty location.

4 - On Equitable Facility Location

Drew Horton, University of Colorado - Denver, Denver, CO, United States, Tom Logan, Daphne Skipper, Emily Speakman

Facility location models traditionally focus on optimizing logistical and cost-related factors and have often overlooked the crucial aspect of equity. However, there are several compelling reasons for integrating equity into facility location models; these include promoting social justice and increased urban resilience. By addressing disparities in access to essential services and resources, facility location decisions can benefit all members of society, particularly marginalized and underserved communities. Furthermore, reducing inequity enhances the long-term sustainability of communities by fostering social cohesion and inclusive economic growth. In this talk, we present recent results that address the question of how best to model equity in various instances of this important domain.

5 - Enhancing Spatial Access to Healthcare: Integrating Location Analytics with Optimization Modeling

Zhiyuan Wei, California Polytechnic State University, San Luis Obispo, CA, United States

Improving spatial access to healthcare facilities is of great interest in urban planning. However, most facility location models fall short of explicitly incorporating spatial accessibility measures when designing facility locations. Thus, we aim to develop a novel analytical framework that directly improves spatial access to healthcare facilities and minimizes access disparities across different population groups. Specifically, a set of machine learning models are trained to estimate healthcare demands from large-scale mobile phone data. Based on human mobility patterns, the travel time decay effect is calibrated based on real road networks. With all these as inputs, a variant of the facility location model is developed to tradeoff between efficiency and equity criteria in enhancing spatial accessibility. To demonstrate the applicability of the proposed framework, a case study is presented. This study could help urban planners determine the optimal locations to establish new hospitals, with the hope of improving equitable access to healthcare services among various socioeconomic groups.

TD45

Summit - 437

Learning and Optimizing for Healthcare Operations

Invited Session

Health Applications Society

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, 19104, United States

1 - Identifying Risk of Intimate Partner Violence with Multi-Modal Machine Learning

Jiayi Gu, Massachusetts Institute of Technology, Cambridge, MA, United States, kimberly villalobos, Yu Ma, Dimitris Bertsimas

Intimate partner violence (IPV) refers to the abuse and aggression from previous or current partners. It is a widespread but underreported public health concern that impacts more than a third of women in the United States. Aside from potentially life-threatening injuries, IPV has a wide range of negative effects on the mental and physical well-being of individuals. This work presents multimodal machine learning models for the early detection of IPV in clinical settings, developed with a dataset of female patients who sought help and support at a violence prevention center of a major hospital in the United States. Utilizing both tabular clinical data and unstructured clinical notes, our models can identify IPV risk with an AUC of 0.88 and years before patients seek help. The models achieved consistent performance in identifying IPV risk for patients who did not seek help at the violence prevention center and patients from another hospital in the same integrated network. This work demonstrates the potential of utilizing multimodal machine learning and historical longitudinal data for the early identification of IPV risk as we work towards developing an IPV risk assessment tool for healthcare settings.

2 - Randomized Controlled Trials for Capacity-Constrained Interventions

Hannah Li, Columbia, New York, NY, United States, Justin Boutilier, Jonas Jonasson, Erez Yoeli

Randomized controlled trials (RCTs), or experiments, are the gold standard for intervention evaluation. However, the main appeal of RCTs---the clean identification of causal effects---can be compromised by interference, when one subject's actions can influence another subject's behavior or outcomes. In this paper, we formalize and study a type of interference in a subclass of interventions we term Service Interventions (SIs): interventions that include an on-demand service component provided by a costly and limited resource (e.g., healthcare providers or teachers). We show that capacity constraints, induced by staffing decisions made by the experimenter, can mediate the effect size of the intervention.

We first show that in such a system, the capacity constraints induce dependencies across experiment subjects, where an individual may need to wait before receiving the intervention. By modeling these dependencies using a queueing system, we show how increasing the number of subjects without increasing the capacity of the system can result in a smaller treatment effect size. This has implications for conventional power analysis: increasing the sample size of an RCT without appropriately expanding capacity can decrease the study's power. To address this issue, we propose a method to jointly select the system capacity and number of users using the square root staffing rule from queueing theory. We show how incorporating knowledge of the queueing structure can help an experimenter reduce the amount of capacity and number of subjects required while still maintaining high power.

3 - Optimizing Health Supply Chains with Decision-Aware Machine Learning

Tsai-Hsuan Chung, Wharton School, Philadelphia, PA, United States, Jatu Abdulai, Patrick Bayoh, Lawrence Sandi, Francis Smart, Hamsa Bastani, Osbert Bastani

This paper combines machine learning (to predict demand) with optimization (to optimize allocations) to address the problem of allocating limited supply of medical resources in developing countries, in particular, Sierra Leone. A key challenge is the need to align the loss function used to train the machine learning model with the decision loss associated with the downstream optimization problem. Traditional solutions have limited flexibility in the model architecture and scale poorly to large datasets. We propose a decision-aware learning algorithm that uses a novel Taylor expansion of the optimal decision loss to derive the machine learning loss. Importantly, our approach only requires a simple re-weighting of the training data, ensuring it is both flexible and scalable, e.g., we incorporate it into a random forest trained using a multitask learning framework. In collaboration with the Sierra Leone government, we deployed our framework in a staggered rollout across all 1,123 government healthcare facilities nationwide. We use synthetic differences-in-differences to evaluate the impact of our tool, finding a more than 20% increase of overall consumption of essential medicines, thereby significantly improving real-world patient access to care.

4 - Modeling and Scheduling Queues with Heterogeneous Resource Demand in Multi-Class Multi-Pool Systems

Weiye Xi, Columbia Business School, New York, NY, United States, Carri Chan, Jing Dong

We study the scheduling of a new class of multi-class, multi-pool queueing systems characterized by heterogeneous resource demands across customer classes. We introduce a model that incorporates multiple types of resources, some with strict capacity constraints and others with soft constraints where exceeding capacity is penalized. These constraints exert different influences on system performance, posing challenges in deriving the optimal policy. Our primary focus lies in the application of this model to hospital inpatient care, where patients require both bed occupancy and specific nursing capacities for adequate treatment. We propose an index-based scheduling policy that aims to optimize the trade-off between holding costs, policy-induced idleness, and constraint violation penalties. We provide insights into the determination of the optimal penalty coefficient based on different levels of system occupancy and its empirical significance. We demonstrate the superior performance of our policy over reasonable benchmarks in several representative scenarios.

TD46

Summit - 438

Fair Decision-making in Healthcare

Flash Session

Health Applications Society

Chair: Wesley Marrero, Thayer School of Engineering at Dartmouth, Hanover, United States

Co-Chair: Zequn Chen, Dartmouth College, Hanover, NH, United States

1 - Fair Distributional Reinforcement Learning for Mental Intervention

Zequn Chen, Dartmouth College, Hanover, NH, United States, Wesley Marrero

The fair distributional reinforcement learning learns the fair return function. We add fair constriant under the classical structure of distributional reinforcement learning and analyze the new algorithm's statistical properties. The new architecture fulfills fairness by guaranteeing similarity of return among different subgroups.

2 - Fairness with Probabilistic Independence

Anil Aswani, UC Berkeley, Berkeley, CA, United States, Mahbod Olfat, Yoon Lee, Ilgin Dogan, Zuo-Jun Shen

There is growing interest on ensuring fairness in algorithmic decision-making. One of the challenges is that the relevant notion or definition of fairness varies based on the underlying context. This talk presents how a definition of fairness based on probabilistic independence can be incorporated into data-driven decision-making based on the framework of statistical decision problems and on the framework of principal-agent models. Examples of applications to healthcare are provided.

3 - Fairly Predicting Graft Failure in Liver Transplant for Organ Assigning

Na Zou, University of Houston, Houston, TX, United States

Liver transplant is an essential therapy performed for severe liver diseases. The fact of scarce liver resources makes the organ assigning crucial. Model for End-stage Liver Disease (MELD) score is a widely adopted criterion when making organ distribution decisions. However, it ignores post-transplant outcomes and organ/donor features. These limitations motivate the emergence of machine learning (ML) models. Unfortunately, ML models could be unfair and trigger bias against certain groups of people. To tackle this problem, this talk will discuss a fair machine learning framework targeting graft failure prediction in liver transplant. Specifically, knowledge distillation is employed to handle dense and sparse features by combining the advantages of tree models and neural networks. A two-step debiasing method is tailored for this framework to enhance fairness. Experiments are conducted to analyze unfairness issues in existing models and demonstrate the superiority of our method in both prediction and fairness performance.

4 - Stacking Machine Learning Models that Incorporate Genetics and Ancestry to Improve Phenotype Prediction: A Case Study

Nahian Tahmin, University of Tennessee, Knoxville, TN, United States, Lokesh K lchintha, Akram Mohammed, Vincenza Colonna, Tesfaye Mersha, Robert L Davis, Anahita Khojandi

This study develops a machine learning (ML) model to explore the genetics and ancestry's role in ICS response in severe asthma (SA). Asthma's progression to SA, which is unresponsive to inhaled corticosteroid (ICS) treatment, represents a significant healthcare burden despite constituting less than 10% of cases. Genetic and ancestry factors play a critical role in SA, but current risk assessment methods do not combine these factors effectively lacking comprehensiveness. Using the Biorepository and Integrative Genomics (BIG) Initiative dataset, we focus on a self-reported African-American pediatric cohort (cases=controls=125) and apply ML to whole exome sequenced SNPs data to address this gap. Data preprocessing includes linkage disequilibrium pruning, Hardy-Weinberg equilibrium filtering, and minor allele frequency filtering. Ordinal encoding marks allele frequency, and missing values are imputed using k-nearest neighbors. Local ancestry data is obtained via RFMix v2 with phased genomic data and a reference panel from the '1000 Genomes Project.' We utilize stratified 10-fold cross-validation for dimensionality reduction and ML modeling, designing two primary pipelines (P1-P2), each with two sub-pipelines. Pipeline 3 (P3) stacks these sub-pipelines into three sub-pipelines (SNP; LA; SNP-LA). P1 applies Lasso regression to SNP and ancestry data, P2 employs Recursive Feature Elimination and Elastic Net followed by Random Forest on SNP and ancestry data. P3 integrates these pipelines with weighted probabilities for a comprehensive model. Results show that incorporating local ancestry data improves ROC AUC from 0.693 ± 0.066 to 0.727 ± 0.041 ($p=0.020$). Both SNP and local ancestry data provide valuable, non-overlapping insights, enhancing genetic prediction accuracy in SA.

5 - Optimizing Vaccination Campaign Strategies Considering Societal Characteristics

Zelda Zabinsky, University of Washington, Seattle, WA, United States, Serin Lee, Shan Liu

Vaccine hesitancy remains a persistent public health challenge. This study examines the interplay between disease transmission, evolving vaccination opinions, and targeted vaccination campaigns. We employ an evolutionary game-theoretic framework to model how individuals make vaccination decisions and how word-of-mouth campaigns influence vaccination uptake. Our findings highlight that vaccination campaign effectiveness and allocation strategies depend on the specific social context. This includes factors like the levels of perceived vaccine risks and the relative weight individuals give to logic versus emotion when making vaccination decisions. In addition to maximizing vaccination uptake, we consider maximizing the minimum vaccination rate increase across demographic groups to account for fairness. Resource allocation for an optimal campaign strategy specifies demographic groups to target. A campaign is most effective when the society has moderate levels of perceived risk of vaccination, and where individuals balance emotional and rational decision-making factors. When the society has low perceived risk of vaccination, natural uptake is already high, making a campaign unnecessary. When the objective is to maximize the minimum vaccination rate increase among groups, the total vaccinated population decreases by 16% compared to the base case. Although overall campaign effectiveness decreases, this approach ensures a more equitable distribution of campaign benefits. This research underscores the importance of considering societal dynamics and demographics when tailoring messaging and targeting groups to maximize the impact of vaccination campaigns.

TD47

Summit - 439

Sequential Decision Making in Healthcare

Invited Session

Health Applications Society

Chair: Hossein Piri, Haskayen School of Business-University of Calgary, Calgary, AB, T2R1C3, Canada

1 - Optimizing Drug Interaction Alerts: A Grouping-Based Approach to Mitigate Alert Fatigue Among Clinicians

Hossein Piri, Haskayen School of Business-University of Calgary, Calgary, AB, Canada, Michael Lingzhi Li

Computerized Physician Order Entry (CPOE) systems play a crucial role in healthcare, but the excessive generation of alerts often leads to alert fatigue among clinicians. To address this challenge, this research proposes a novel approach that utilizes artificial intelligence (AI) to categorize alerts based on their severity. By employing a Partially Observable Markov Decision Process, the research aims to strategically select and present alerts to prescribers, thereby reducing the likelihood of alert fatigue. This grouping-based methodology seeks to optimize the alert system by intelligently filtering out less critical alerts and prioritizing those of higher importance. Through the utilization of AI-driven severity classification and decision-making processes, the study endeavors to enhance the effectiveness of alert notifications within CPOE systems. By focusing on presenting alerts that are most relevant and crucial for clinical decision-making, this research strives to improve the overall alert management process and alleviate the burden of excessive alerts on healthcare providers. The implementation of this approach has the potential to streamline alert delivery, enhance provider acceptance of alerts, and ultimately contribute to more efficient and effective clinical workflows.

2 - Towards Interpretable Sequential Machine Learning

Xiaolong Luo, Harvard University, Allston, MA, United States, Michael Lingzhi Li, Scott Wang, Scott Wang

We introduce the Sequential Multiple Instance Learning (SMIL) framework, addressing the challenge of interpreting sequential, variable-length sequences of medical images with a single diagnostic label. Diverging from traditional MIL approaches that treat image sequences as unordered sets, SMIL systematically integrates the sequential nature of clinical imaging. We develop a bidirectional Transformer architecture, BiSMIL, that optimizes for both early and final prediction accuracies through a novel training procedure to balance diagnostic accuracy with operational efficiency. We evaluate BiSMIL on three medical image datasets to demonstrate that it simultaneously achieves state-of-the-art final accuracy and superior performance in early prediction accuracy, requiring 30-50\% fewer images for a similar level of performance compared to existing models. Additionally, we introduce SMILU, an interpretable uncertainty metric that outperforms traditional metrics in identifying challenging instances.

3 - Guiding Physicians with Time-dependent Patient Selection Policies Under Shift Work

Marco Bijvank, University of Calgary, Calgary, AB, Canada, Mahdi Shakeri

Patient selection in healthcare settings with walk-in patients is often a decision at a physician's discretion. Especially in overcrowded environments such as Emergency Departments (EDs), this can significantly impact the operational performance of a healthcare unit and the clinical outcomes for patients. We propose a time-dependent policy to guide physicians in their patient selection decision, where we consider new patients waiting for an initial assessment, returning patients waiting for a re-assessment, and potential patient transfers between physicians at the end of a physician's shift (also called hand-offs). We formulate a transient optimal control problem and demonstrate that a time-threshold-type policy is optimal. This patient selection policy prescibes which patient type (new or returning) and which patient class (urgent or non-urgent) to prioritize. Furthermore, the actual time thresholds when to switch priorities depend on the status of the ED when the physician starts their shift. Under special cases we are able to derive easy-to-implement selection priority rules (that can be seen as extensions of the well-known cμ-rule). Simulation results based on a case study indicate that our proposed policy significantly reduces patient hand-offs compared to alternative policies from the literature without compromising the length of stay (LOS), patient wait times, or the number of patients who leave without being seen by a physician (LWBS).

TD48

Summit - 440

Adaptive Experimental Design in Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: David Simchi-Levi, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Adaptive Experimental Design in Operations

Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, David Simchi-Levi

Experimentation serves as a cornerstone across diverse fields, from statistics to social science, enabling us to validate theories and optimize strategies. Traditional experimental designs have been largely rooted in statistical objectives—primarily aimed at minimizing bias and maximizing statistical power. Yet, when operational objectives come into play, these conventional methods encounter distinct challenges. This talk will articulate such challenges and explore how adaptive design strategies can provide solutions.

Firstly, we confront the challenge of welfare loss inherent in traditional randomized experiments. By approaching experimental design as a minimax multi-objective optimization problem, we expose the inherent trade-off between in-experiment welfare and statistical power. We introduce the concept of Pareto optimality to mathematically characterize the situation in which neither the statistical power nor the welfare efficiency can be improved without degrading the other. We derive a useful sufficient and necessary condition for the Pareto optimal solutions to the minimax multi-objective optimization problem. Additionally, we design an effective Pareto optimal multi-armed bandit experiment that can be tailored to different levels of the trade-off between the two objectives. Moreover, we reveal an optimal way to design price experiments, where the price is a continuous decision variable. Extending beyond the price elasticity estimation, we are also interested in maximizing the expected revenue through the experiment, and controlling for the tail risk that, if not controlled, may lead to significant financial losses. Finally, we challenge posed by non-stationarity, meaning that the effects of interventions may evolve over time. Neglecting this dynamic can lead to flawed decisions.

2 - Exploring Drug Candidates: All Epsilon-Best Arms Identification in Linear Bandits

Tianyi Ma, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhekai Li, Cheng Hua, Ruihao Zhu

Motivated by the needs to successfully identify multiple candidates that can dramatically enhance outcomes in complex and high-stakes tasks such as drug discovery, we propose a nearly optimal adaptive allocation policy to identify all epsilon-best arms. Specifically, we introduce LinFACTE, an algorithm designed to optimize the identification of all epsilon-best arms in linear bandits. To the best of our knowledge, we provide the first information-theoretic lower bound on the complexity of this problem. Additionally, we demonstrate that our algorithm achieves instance optimality, matching this lower bound up to a logarithmic term. Numerical results demonstrate the practical advantages of LinFACTE compared to baseline methods, highlighting its ability to accelerate early-stage drug development.

3 - Rebandit: Personalizing Treatment Delivery for Reducing Cannabis use

Susobhan Ghosh, Harvard University, Cambridge, MA, United States, Yongyi Guo, Pei-Yao Hung, Lara Coughlin, Erin Bonar, Inbal Nahum-Shani, Maureen Walton, Susan Murphy

As the world moves from reactive care to preventative care, digital health interventions offer an unprecedented opportunity to promote healthy behaviors at scale. However, these interventions often suffer from low engagement and adherence over time. Personalizing the delivery of interventions based on each user's evolving context and preferences can enhance engagement and efficacy. Personalizing treatment in digital intervention studies becomes challenging due to noise and short time horizons. To address this, we present reBandit, an online reinforcement learning algorithm currently utilized in the MiWaves mobile health study. MiWaves aims to deliver personalized interventions to reduce cannabis use among emerging adults. reBandit leverages population heterogeneity by utilizing random-effects, and uses informative Bayesian priors, both enabling efficient learning in noisy mobile health settings. Furthermore, reBandit employs Empirical Bayes and optimization techniques to autonomously update its hyper-parameters online. We benchmark the performance of our algorithm using a simulation testbed. We show that reBandit performs equally well or better than all baseline algorithms, with its performance increasing as population heterogeneity increases.

4 - Planning Adaptive Experiments: A Mathematical Programming Approach

Jimmy Wang, Columbia University, New York, NY, United States, Daniel Jiang, Ethan Che, Hongseok Namkoong

Adaptive experimentation can improve statistical power significantly, but typical algorithms overlook important issues that arise in practice: multiple objectives, non-stationarity, batched/delayed feedback, constraints, and personalization. Moving away from developing bespoke algorithms for each setting, we present a mathematical programming view of adaptive experimentation that can flexibly incorporate a wide range of objectives, constraints, and statistical procedures. By formulating a dynamic program in the batched limit, our modeling framework enables the use of scalable optimization methods (e.g., SGD and auto-differentiation) to solve for treatment allocations. To spur algorithmic progress, we build a suite of benchmark problems based on hundreds of real A/B tests at ASOS that model key practical issues such as non-stationarity, personalization, multi-objectives, and constraints. Our empirical results show standard Thompson sampling-based policies fail to reliably improve upon static designs, and demonstrate the effectiveness of a simple planning approach.

5 - AI-Assisted Fashion Design: LLM and Diffusion Models for Improving Market Demand

Junyi Sha, Massachusetts Institute of Technology, Cambridge, MA, United States, Xiaomin Li, David Simchi-Levi, Michelle Wu

Identifying the key features that influence consumer preferences is crucial in the fashion industry. In this study, we develop a comprehensive methodology to pinpoint the most impactful features of fashion products based on actual market sales data. We first introduce a defined scoring system for product features. Then, utilizing diffusion models, we modify product images to isolate and assess specific attributes. Subsequently, we employ machine learning algorithms—random forest and XGBoost—along with evaluations from human raters to verify the performance of AI-modified products. Through this combined approach, we demonstrate that enhancing the identified "good" features leads to significant improvements in product market performance. Our findings suggest that integrating machine learning techniques with human judgment provides a powerful tool for optimizing fashion product design and marketing strategies.

TD49

Summit - 441

Data Collection and Analytics

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Lennart Baardman, University of Michigan, Ann Arbor, MI, United States

1 - Managing Product Availability in Rental Subscriptions

Xuening Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Lennart Baardman

The emergence of rental subscriptions in industries like clothing, furniture, and transportation provides customers access to products that may not be cost-effective to purchase outright. However, rental subscription businesses face significant challenges in reducing customer churn. In collaboration with a clothing rental firm, we identified insufficient product variety as a frequent cause of churn. To address this, we developed fast and implementable policies for the underlying product display and fulfillment dynamic optimization problem. Our prediction-based policy forecasts inventory availability for future rental dates, while the scheduling-based policy makes decisions based on a nominal schedule. Testing these policies on real-world, calibrated simulations, we achieved up to a 5% reduction in overall churn rate. Additionally, we provided theoretical performance bounds for these strategies, comparing them to the hindsight optimal solution and the current instant fulfillment approach.

2 - Costly Data Collection in the Newsvendor Problem

Zijin Zhang, University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Hyun-Soo Ahn, Lennart Baardman

In recent years, data-driven algorithms and policies have flourished in many areas of operations management. While some applications can utilize the preexisting data, there are many cases where collecting plenty of high-quality data is often cost-prohibitive and time-consuming. Previous research has mainly focused on improving decision-making with given data, leaving a gap in understanding the optimal quantity and quality of data needed. Our paper studies how the the sample size (quantity) and noise scale (quality) of a dataset jointly affects the outcome of data-driven decision making in the newsvendor problem. To mitigate the effect of noise, we propose a novel denoised newsvendor policy that minimizes the noise-robust in-sample cost given a noisy dataset. Using the denoised approach, we analyze the retailer's out-of-sample newsvendor outcome as a function of sample size and noise scale. Our analysis quantifies the marginal newsvendor cost on each dimension, and we find that the newsvendor outcome is non-monotone in neither quantity nor quality dimension. Furthermore, given the cost of data collection, we develop a series of data collection policies and show that collecting a dataset of well-balanced quantity and quality is important.

3 - Exploiting Human and AI Synergies: An End-to-End Selective Classification Framework

Bradley Rava, University of Sydney Business School, Sydney, Australia, Michael Huang

Modern Artificial Intelligence (AI) tools have reached a point where policymakers can potentially engage in fast and autonomous large-scale decision-making in high-risk settings. However, despite advancements, the adoption of AI in these settings remains controversial as they can be inaccurate and disproportionally affect minority or underrepresented groups and individuals. As a result, many policymakers revert to human-based decision-making, which, unfortunately, can be equally ineffective as decisions are often delayed or require high labor costs. To demonstrate this issue, we investigate the New York Department of Transportation's speed reducer tracking system. The data from this system can inform decisions regarding the placement of speed humps within the five boroughs. Our presentation introduces an end-to-end selective classification approach that optimally divides decision-making between human expertise and artificial intelligence. The proposed method aims to automatically recommend speed hump placements, while ensuring fairness across the different boroughs. Additionally, it leverages human judgement for decisions where human input is most valuable.

4 - Nuancing majority vote recommendations with Large Language Models

Oliver Schaer, Drexel University, Philadelphia, PA, United States, Matthew Schneider, Nikolaos Kourentzes, Panos Markou

Expert committees are widely used to support decision-makers on complex issues. In many cases, these expert groups submit their recommendations via a majority vote. Although efficient, relying on voting information alone loses experts' discussion, including important nuance, concerns, or uncertainty. For example, a member might vote in favor of a question but add a caveat or other important context that does not make it to the decision table where votes are presented. We suggest Large Language Models to efficiently capture those expressed nuances. More specifically, we want to take advantage of the stochastic nature of LLMs to obtain distributions around the voters' perceived uncertainty level. We empirically evaluate the proposed methodology on FDA Advisory Committee meeting transcripts where the board evaluates novel drugs and medical devices.

TD50

Summit - 442

MSOM: Analytics for Social Goods

Invited Session

MSOM: Healthcare

Chair: Pengyi Shi, Purdue University, West Lafayette, IN, 47907, United States

Co-Chair: Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Optimizing Access to Primary Care

Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States

We propose a joint predictive-prescriptive framework for primary care transformation. Primary care departments traditionally take a "one size fits all" approach to panel assignment and care delivery, which can lead to over-burdened providers and long delays for booking appointments. However, the heterogeneity of patients within a panel, and the emergence of more flexible care models, provide an opportunity to transform department operations. We develop models to predict individual patient demand and an optimization framework for panel assignment. Our model accommodates multiple provider levels (e.g., MDs and APRNs) and visit modalities (in person vs. virtual), providing a tool to evaluate the tradeoffs between current practice and alternative models. The model is currently in use to inform department care delivery strategies across a large CT health system.

2 - Improving the Security of United States Elections with Robust Optimization

Bradley Sturt, University of Illinois Chicago, Chicago, IL, United States, Braden Crimmins, Alex Halderman

For more than a century, election officials across the United States have inspected voting machines before elections using a procedure called Logic and Accuracy Testing (LAT). This procedure consists of election officials casting a test deck of ballots into each voting machine and confirming the machine produces the expected vote total for each candidate.

In this talk, we bring a scientific perspective to LAT by introducing the first formal approach to designing test decks with rigorous security guarantees. Specifically, we propose using robust optimization to find test decks that are guaranteed to detect any voting machine misconfiguration that would cause votes to be swapped across candidates. Out of all the test decks with this security guarantee, the robust optimization problem yields the test deck with the minimum number of ballots, thereby minimizing implementation costs for election officials. To facilitate deployment at scale, we developed a practical exact algorithm for solving our robust optimization problems based on mixed-integer optimization and the cutting plane method.

In partnership with the Michigan Bureau of Elections, we retrospectively applied our robust optimization approach to all 6928 ballot styles from Michigan's November 2022 general election; this retrospective study reveals that the test decks with rigorous security guarantees obtained by our approach require, on average, only 1.2% more ballots than current practice. Our robust optimization approach has since been piloted in real-world elections by the Michigan Bureau of Elections as a low-cost way to improve election security and increase public trust in democratic institutions.

3 - Learning Innovation Curves

Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States, Christian Kaps

In sectors such as healthcare and renewable energy how should investors make decisions about which new technologies to invest in? The development of these new technologies can be described using experience curves, also known as innovation curves or learning curves. These allow us model the decrease in cost of a technology as a function of the cumulative investment into it. It is important for investors to balance the tradeoff between investing into multiple different technologies and learning to better estimate the structure of these learning curves, versus dedicating resources to the technology that seems to be the best at the current decision point. We model this decision-making problem as a dynamic program, analyze the performance of different investment strategies, and offer insights into the conditions under which certain strategies might outperform others.

4 - Latent Feature Mining via LLMs: An Application in Incarceration Diversion Program

Pengyi Shi, Purdue University, West Lafayette, IN, United States, Bingxuan Li, Amy Ward

Decision making often operate under constraints of incomplete or sparse datasets with missing entries. Recent advancements in large language models (LLMs) offer a promising avenue to bridge these data gaps. In this work, we propose framework to leverages LLMs to augment observed features collected in given datasets with latent features, enhancing the predictive power of ML models for downstream tasks, and supporting the decision-making. We use data from the criminal justice setting to demonstrate the efficacy of this framework.

TD51

Summit - 443

Unveiling Complexity: Simple Algorithms for Complex Systems

Invited Session

MSOM: Service Operations

Chair: Xu Sun, University of Miami, Miami, FL, United States

1 - Dynamic Control of Multiclass Queueing Networks in Heavy Traffic: Interpreting Solutions to Brownian Approximations

Baris Ata, University of Chicago, Booth School of Business, Chicago, IL, United States, Michael Harrison, NIAN SI

We investigate Brownian approximations for control problems in queueing networks, specifically addressing the complex scenario where the equivalent workload formulation is multi-dimensional. By employing deep learning, we solve the approximating Brownian control problem and propose a method to implement this solution in the original queueing network problem. We demonstrate the effectiveness of our method through performance comparisons with benchmarks using simulation.

2 - Real-TIME Spatial Intertemporal Pricing and Relocation in a Ride-Hailing Network

Chen Qi (George), London Business School, London, United Kingdom, Murray Lei, Stefanus Jasin

This paper studies a tactical spatial–intertemporal pricing problem where a firm uses a pool of homogeneous servers to serve price-sensitive customers within a finite horizon. We develop a simple state-dependent pricing policy and establish its asymptotic revenue loss performance which strictly improves over state-independent pricing policies. Similar insights hold when the firm is also allowed to relocate servers at a cost in addition to pricing.

3 - Cloud Cost Optimization: Model, Bounds, and Asymptotics

Zihao Qu, The University of Massachusetts Amherst, Amherst, MA, United States, Milind Dawande, Ganesh Janakiraman

Motivated by the rapid growth of the Cloud Cost Management and Optimization (CCMO) industry to support the exploding cloud-computing market, we study an infinite-horizon, stochastic optimization problem from the viewpoint of a firm that employs cloud resources to process incoming jobs over time. We model the following practical features of CCMO in our problem. There are several types of resources differing in their costs and performance attributes (e.g., processor speed, memory). For each type of resource, capacity can either be reserved over the long term at a discounted price or used on-demand at a relatively higher price. Jobs of several types arrive stochastically through time -- jobs differ in their completion-time deadlines and in their resource-specific processing-time distributions. Moreover, the progress of a job can be observed periodically and, if required, the job can be moved from one resource-type to another. Penalty costs are incurred for jobs not completed by their deadlines. The firm's goal is to minimize the long-run average expected cost per period, considering reserved-capacity costs, on-demand capacity costs, and job-delay costs. We derive a lower bound on the optimal cost by considering a set of decoupled problems, one for each job. The solutions of these problems are then used to construct a feasible policy for the original problem and derive an upper bound on that policy's optimality gap. Importantly, we show that our policy is asymptotically optimal. We also report results of a comprehensive numerical study on AWS data to demonstrate the impressive performance of our policy.

4 - Asymptotically Optimal Dispatch Policies for Emergency Medical Services

Jingwei Zhang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Cheng Hua, Tong Wang, Ziyan Zhou

Emergency medical services (EMS) play a critical role in ensuring timely and effective healthcare delivery, particularly in situations where rapid response is important. The dispatching of ambulances, a fundamental component of EMS, is a complex decision-making process that directly influences patient outcomes. Cities are faced with the challenge of managing limited resources to respond promptly to current emergency calls while maintaining the capacity to address potential future services. In this paper, we consider the dynamic nature of call arrivals and the inherent heterogeneity in service times across different station-location pairs with general distributions. Our goal is to determine which unit to dispatch to each sequentially observed call to minimize the system-wide average cost. To address this problem, we develop two easy-to-implement and near-optimal policies based on a deterministic linear program and a Lagrangian relaxation. Our case study, based on real data from New York City, demonstrates how implementing various proposed policies can effectively reduce costs under different levels of demand.

5 - Optimal Design and Re-Design of Large Scale Service Systems: Pricing and Capacity Optimization with Price and Congestion Sensitive Customers

Levent Kocaga, Sy Syms School of Business, New York, NY, United States

We study the joint pricing and capacity optimization and re-optimization problem faced by a service provider that is modeled as an unobservable queue. Customers are price and delay sensitive with heterogeneous service valuations which subsequently determine the endogenous arrival rate via a unique equilibrium for a given price and capacity. The objective of the service provider is to choose the price and/or capacity that maximizes the average profit rate in equilibrium. Since the exact analysis of such a system is only possible numerically, we pursue an asymptotic analysis by letting the market size grow large while holding parameters that define individual customer characteristics fixed. Our main theoretical contribution is to show that, under various formulations, the optimal system operates under the Quality-Efficiency-Driven (also known as the Halfin-Whitt) heavy traffic regime. An important practical implication of the QED regime is that we can derive tractable analytical expressions for the optimal price, capacity, and profit rate which enable us to glean novel structural insights that are relevant to the operation of large-scale service systems (and are otherwise not possible via exact analysis). Our results can also help guide service providers to run promotional marketing activities that increase market size such as paid advertising or free trials.

TD52

Summit - 444

Social Responsibility and Sustainability

Invited Session

MSOM: Service Operations

Chair: Serguei Netessine, The Wharton School, Philadelphia, PA, United States

1 - Optimizing Impact of Solar Lighting on the Lives of Poor: Evidence from Field Experiments in Ghana

Bhavani Shanker Uppari, Singapore Management University, Singapore, Singapore, Serguei Netessine

We randomly assigned solar-lighting kits with varying configurations to a large number of households in Ghanaian villages. We document the impact of these kits on these treatment households, in terms of economic improvement as well as improvements in the education of children and health of the household members, vis-a-vis control households who were not given solar kits. We also build a structural model that allows us to optimize for the impact in a fast and cost-effective manner.

2 - Harvest Picking in the Era of Global Warming: Hire Now or Later?

Masha Shunko, University of Washington, Seattle, WA, United States, Sanchita Das, Leonard Boussioux

In collaboration with the Washington Fruit Tree Research Center, we design a capacity planning tool that enables dynamic resource allocation for fruit tree growers in response to climate-based yield uncertainties. Given the fluctuations in weather and climate due to global warming, the harvest windows for multiple crops are increasingly more variable over the years. We develop a robust optimization technique that dynamically allocates resources prior to harvesting season.

3 - Beyond Cookies: Evidence about Team Environment and Engagement Retention from Girl Scouts Cookie Program

Tom Tan, Southern Methodist University, Dallas, TX, United States, Bradley Staats

The learning process requires the participation from both the individuals and their teammates, creating a “coproductive” process. Although past research has emphasized various impacts of team coproduction factors, few studies have been done to understand the engagement retention effect, and its impact on learning outcome, particularly in a nonprofit setting.

We focus on this understudied area and study five coproduction factors in the setting of Girl Scout Cookie Program. They include troop size, the disparity of sales performance within the troop, the ratio between adult and girl members, the grade level variety, and troop-sponsored booth sales.

We find that a small troop size and more evenly distributed sales performance within the troop are positively associated with the propensity to return to the Cookie Program. A high adult-to-girl ratio and a significant grade level (i.e., age) diversity in the troop are also conducive to scout retention. Furthermore, conducting troop-based booth sales encourages more scouts to return. We further find that a more evenly distributed sales performance within the troop increases the likelihood to return for both high- and low-sales-performance Girl Scouts, but the positive effect is stronger for high-performance ones than low- performance ones. By contrast, the large grade level variety increases the retention for both older and younger girls. Finally, we find evidence to support that the retained engagement and its associated "learning-by-doing" experience indeed enhance the learning outcome, and that almost all the five team coproductive factors, except troop size, directly enhance learning to sell, too.

4 - Hyperlocal P2P Food Sharing: Estimating the Effect of the Entry of the Commercial Food Donation Channel

Ekaterina Astashkina, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Masha Shunko, Yixin Iris Wang

We are examining data from the largest global peer-to-peer food sharing platform. This platform facilitates the exchange of soon-to-expire food items among platform's users. We investigate the impact of introducing a new commercial donation channel, which allows various food establishments (e.g., supermarkets, grocery stores, bakeries) to donate surplus food directly to households. Specifically, we aim to understand how this addition influences donation patterns among users of the platform.

TD53

Summit - 445

OM Research on Uplifting Low-income Communities

Invited Session

MSOM: Sustainable Operations

Chair: Wei Wei, University of New Hampshire, Durham, NH, United States

Co-Chair: Priyank Arora, University of South Carolina, Columbia, SC, 29208, United States

1 - Subsidizing Social Welfare Programs: Contracted Slots OR Vouchers?

Wei Wei, University of New Hampshire, Durham, NH, United States, Priyank Arora, Senay Solak

We study operational decision-making by profit-oriented service providers under two popular service-focused subsidy welfare programs—namely, vouchers and contracted slots. We analyze how program-, provider-, and market-related factors influence the quantity and quality of essential services provided by the providers. Through the comparison of resulting equilibrium outcomes under these two program types, we identify conditions under which it is beneficial (in terms of societal impact generated per cost incurred) for the government to rely on vouchers versus contracted slots.

2 - Improving Process Flow in a Child-Welfare System: A Data-Driven Approach

Shubham Akshat, Carnegie Mellon University, Pittsburgh, PA, United States, Alan Scheller-Wolf, Justin Galbraith, Lindsey Lacey

We study the problem of a case manager of a Child Welfare Organization (CWO).They receive suspected cases of child maltreatment and provide services (includes counseling, support services for independent living, and foster placement) to the child and family.

The CWOs are short-staffed; the true nature of a referral is not known upfront and requires effort (in the form of investigation by a caseworker) to reveal its true type; errors can be very damaging. For example, screening in an innocuous referral for investigation creates negative social value (inconvenience to the child and family) while this effort would have created positive social value if it were spent on investigating a true high-risk referral. Screening out of a high-risk referral carries even greater potential risk. Thus, we have two-sided errors in our setting.

This study aims to find 1. an optimal admission policy (for new referrals) and 2. determine the optimal effort allocation of a caseworker over the set of heterogeneous referrals (in their priority and age) to maximize the throughput and reduce the error rates in determining case/referral resolutions.

Overall, our study lies at the intersection of production queuing network, multi-class queues, and discretionary task completion. We apply OR/OM tool to a child-welfare system and we use real data to estimate policy parameters.

3 - The High Impact of Disasters on Prices in Low-Income Communities

Xabier Barriola, Tilburg University, Tilburg, Netherlands, William Schmidt

Responding to natural disasters requires the deployment of food, shelter, and medical assistance to relieve the immediate needs of the victims. These services are supported by, and eventually fully transitioned to, local resources, such as grocery stores. However, disparities in local infrastructure and services may result in low-income communities bearing a disproportionate burden from these disasters. We investigate this possibility by testing for differences in the percentage change in paid prices, promotions, stockouts, and substitutions for grocery products between low-income and high-income communities following three large Atlantic hurricanes. Using a triple-difference regression specification, we find that low-income communities in the disaster zones endure higher average percentage price increases within grocery categories compared to high-income communities. We establish the presence of several mechanisms that contribute to this outcome -- a larger percentage drop in price promotions, higher percent price increases at the product level, more frequent stockouts, and a larger increase in substitution from low-priced products to high-priced products.

4 - Managing Payment Flexibility in Rent-to-Own Contracts for off-Grid Energy Products

Gonzalo Romero, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Jose Guajardo, Elaheh Rashidinejad, Hosain Zaman

The diffusion of technological innovations in developing economies has been facilitated by the use of Rent-To-Own (RTO) business models, which give flexibility to consumers by allowing them to make incremental payments over time. Motivated by an application of RTO to the distribution of solar lamps in developing countries, we examine the drivers and impact of payment flexibility on repayment performance and consumer behavior in RTO contracts.
We formulate a stochastic dynamic programming model that characterizes an important dimension of payment flexibility, i.e., the ability of consumers to make bundled payments (multiple installments paid at once, in advance). We show that consumers may bundle payments because of uncertainty about budget in future periods, which leads to a non-monotonic impact of income uncertainty on repayment performance. We further show that bundled payments are more likely to occur closer to the end of the ownership cycle. We examine different flexibility levers that the firm can adjust as part of its contract design (repayment frequency/installment amount, grace period), accounting for the impact of bundled payments. Our results show that an intermediate level of flexibility could benefit both the firm and consumers under some conditions.
.

5 - Improving Cash-Constrained Smallholder Farmers' Revenue: the Role of Government Loans

Yanchong Zheng, Massachusetts Institute of Technology, Cambridge, MA, United States, Jimin Park, Somya Singhvi

The need for immediate cash often forces farmers in developing countries to sell their crops when the market price is low, thus hurting their revenue. This paper examines how cash constraints influence farmers' selling decisions across the harvest and lean seasons, and analyzes the efficacy of government loans in improving farmers' revenue. We develop a game-theoretic model to characterize the market equilibrium under the base scenario of no government loan, uncovering the impacts of cash constraints on the revenue of farmers with different production quantities. We then examine how a government loan program may counteract these negative impacts. Under a government loan, farmers store some of their production quantities at government warehouses in exchange for immediate cash. We analyze and contrast two types of loan policies, one in which all farmers are eligible for the loan (a homogeneous loan) and the other in which the loan is only offered to farmers whose production quantity is below a certain threshold (a heterogeneous loan). Our results demonstrate that heterogeneous loans are more desirable when cash constraints are moderately high. We further show that, when designed properly, the loan can simultaneously increase aggregate farmer revenue and generate more equitable revenue distribution among the farmers. Taken together, these results underscore that government loan design must carefully account for farmers' strategic response to the policy to generate positive societal outcomes. Finally, we use field data of Bengal gram farmers in India to empirically validate our insights and quantify the revenue impact of the policies studied.

TD54

Summit - 446

Sustainability and Business Model Innovation

Invited Session

MSOM: Sustainable Operations

Chair: Elena Belavina, Cornell University, New York, NY, 10044, United States

Co-Chair: Yu Nu, Cornell University, New York City, NY, United States

1 - Seeing Beauty in Ugly Produce: Do Ugly Produce Retailers Reduce Food Waste?

Zheng Han, DePaul University, Chicago, IL, United States, Bin Hu, Milind Dawande

Ugly produce refers to cosmetically-unconventional yet taste- and nutrition-wise perfectly edible produce. Advocates of ugly produce retailers claim that they help reduce food waste. Skeptics, however, argue that ugly produce is often not wasted but meets alternative demands such as from restaurants and processed food manufacturers. Given such alternative demands, do ugly produce retailers reduce food waste? We model an agricultural cooperative producing, pricing and selling produce to a retailer and an alternative buyer, and find that the retailer selling ugly produce still strictly reduces food waste. However, doing so also generally reduces the retailer’s profit, which explains traditional grocery chains’ hesitancy to carry ugly produce. Fortunately, we find that a dedicated ugly produce retailer competing with a traditional retailer yields the same food waste reduction. Therefore, whereas traditional retailers could but would not carry ugly produce and help reduce food waste, ugly produce startups can willingly play this role in their stead. This research sheds light on the ongoing debate about whether ugly produce retailers reduce food waste and explains the development of the ugly produce movement. Our findings justify the sustainability value of ugly produce retailers—especially dedicated ugly produce startups—and inform the collaboration strategies of non-profit organizations with farmers and/or retailers to reduce food waste and combat global hunger and climate change.

2 - Using Artificial Intelligence for Reducing Food Waste in Commercial Kitchens

Karan Girotra, Cornell Tech/Johnson Cornell University, New York, NY, United States, Yu Nu, Elena Belavina

In this study, we estimate the reduction in food waste from the deployment of a system that digitally records instances of food items discarded in a commercial kitchen. We also shed light on the mechanisms that drive this impact. In a quasi-experimental setting, where the system was deployed in about 900 kitchens in a staggered manner, we estimate the impact using synthetic difference-in-differences method. We find that three months after adoption, kitchens generate 29% lower food waste, on average, than they would have in absence of the system, without any corresponding reductions in sales. Utilizing a long-short-term-memory fully-convolutional-networks classifier, we document that these reductions are accompanied by a 23% decrease in demand chasing, a known bias in human inventory management. Upgrading to a system that uses computer vision for waste classification leads to a further 30% reduction in food waste generated by the kitchen, a year after the upgrade. This further reduction is due to accurate recording of infrequent but very high-impact instances of food wasted that are often not entered in the the manual classification system. We also observe substantial effect heterogeneity by kitchen size, service type, and demand variability.

3 - Organic Waste Bans Have Failed to Divert Waste Away from Landfills in the United States&Mdash;Except in Massachusetts

Fiorentia Zoi Anglou, University of Texas at Austin, Austin, TX, United States, Robert Sanders, Ioannis Stamatopoulos

Diverting food waste away from landfills is a critical step to reduce emissions and meet emissions reduction goals set by the Paris Agreement. Accordingly, between 2014 and 2024, nine US states banned commercial waste generators—such as grocery chains—from landfilling organic waste, expecting a 10-15% waste reduction. However, no evaluation of these bans exists. We study the bans' effects on landfilled waste by compiling and analyzing a comprehensive waste dataset covering 36 US states between 1996 and 2019, and 89\% of the US population's waste disposal. Contrary to policymakers' expectations, we can reject aggregate waste reductions higher than 3.2%, and we cannot reject a zero-null aggregate effect. Moreover, we cannot reject a zero-null effect for any state in our data, except for Massachusetts, which gradually achieved an 11.2% reduction—matching expectations. Our findings reveal an urgent need to reassess the design and execution of such bans, using Massachusetts as a benchmark for success.

4 - Sustainable Technology in the Lemons Market: the Case of Battery Choices by Electric Vehicle Owners in Bangladesh

Yewon Kim, Stanford University, Stanford, CA, United States, Amrita Kundu, Erica Plambeck

We study how public and private interventions can resolve information asymmetry and value misalignment between customers and manufacturers of new sustainable durable goods in the context of electric vehicle (rickshaw) batteries in Bangladesh. Two challenges prevent the widespread adoption of new sustainable technologies. First, due to limited information available to customers at the entry phase, innovations face greater information asymmetry than existing goods. Second, as new sustainable technologies often appear as bundles of greater performance in the form of longer duration, higher prices, and lower social costs, their value proposition creates potential misalignment with customers' needs for short-term usage of products at lower private costs. Although developing economies are likely to enjoy higher benefits from sustainable technologies due to the negative externalities of existing alternatives, they impose greater challenges on these innovations because of a lack of a credible signaling system ("lemons markets") and low average income that leads to financial constraints and consumer myopia. Running qualitative interviews and field experiments in two cities in Bangladesh, we aim to measure the negative effects of information asymmetry and value misalignment in sustainable durable goods adoption and propose interventions that reduce such frictions. Our findings help sustainable technologies with high social benefits penetrate markets more effectively and offer implications for policymakers aiming at switching to sustainable markets.

TD55

Summit - 447

Disaster Response Operations and Social Responsibility

Invited Session

Public Sector OR

Chair: Zhili Tian, Univ. of Houston, Houston, TX, United States

1 - A Multiperiod Multicommodity Capacitated International Agricultural Trade Network Equilibrium Model with Applications to Ukraine in Wartime

Dana Hassani, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Oleg Nivievskyi, Pavlo Martyshev

The world is facing immense challenges due to increasing strife and the impacts of climate change, with accompanying disasters, both sudden-onset as well as slow-onset ones, which have affected the trade of agricultural commodities needed for food security. In this paper, a multiperiod multicommodity international agricultural trade network equilibrium model is constructed with capacity constraints on the production, transportation, and storage of agricultural commodities. The model allows for multiple routes between supply and demand country markets, different modes of transport, and storage in the producing and consuming countries as well as in the intermediate countries. The generality of the underlying functions, coupled with the capacity constraints, allow for the modeling of competition among agricultural commodities for production, transportation, and storage. The capacity constraints also enable the quantification of various disaster-related disruptions to production, transportation, and storage on the volumes of commodity flows as well as on the prices. A series of numerical examples inspired by the effects of Russia's full-scale invasion of Ukraine on agricultural trade is presented, and the results are analyzed to provide insights into food insecurity issues caused by the war.

2 - Socially Responsible Newsvendor

Chen Hu, Xi'an Jiaotong-Liverpool University, Suzhou, China, People's Republic of, Ming Hu, Yongbo Xiao

We focus on a socially responsible firm that is concerned with its profit as well as consumer surplus and examine four different types of pro-social behavior by the firm: optimizing a weighted average of the expected profit and consumer surplus (referred to as the mixed-objective model), negotiating with pro-social executives or consumer activists (referred to as the Nash bargaining), charitable donations after profit maximization (referred to as the donation), and ensuring the portion of consumer surplus to be a given fraction of the social welfare (referred to as the fairness model). Our results show that under all behaviors, there is a more substantial boost to consumer surplus at the expense of a slight decrease in profit when consumer surplus consideration is lower. Among those four behaviors, while maintaining the same profit level, a donation is not the most consumer-surplus-enhancing pro-social behavior among those four behaviors, when the overhead cost is sufficiently high or when a high enough profit level needs to be maintained. This finding challenges Milton Friedman's advocacy that socially responsible businesses should indirectly fulfill their societal duties by focusing on profit maximization and then redistributing the generated profit for social causes (Friedman 1970). Our results imply and quantify that a little commitment to consumers can go a long way in balancing consumer surplus against profit loss. We also shed light on when the firm should choose a decentralized pro-social behavior, such as donations, and when it should incorporate consumer surplus consideration into operational decisions for consumer surplus enhancement.

3 - Harnessing the Potential of Generative Artificial Intelligence for Ad Hoc Logistical Disaster Response

Frank Schätter, Pforzheim University of Applied Sciences, Pforzheim, Germany, Florian Haas

Methods from Operations Research (OR) and Management Science (MS), applied within Decision Support Systems (DSS), ensure that logistical disaster response is based on analytically sound decisions, e.g., by using optimization models for resource allocation, facility location, or transportation planning. To be used for immediate disaster response, the system itself must be operational. This is a challenge when integrating OR/MS methodologies into a DSS, which are time-consuming to implement because they rely on ad hoc information such as environmental impacts and planning conditions. Making assumptions about all these contingencies in the pre-disaster phase is not targeted, as it is impossible to know in advance what will happen, when it will happen, and what the framework conditions will be. This research builds on our previous work developing ReDRiSS (Reactive Disaster and Supply Chain Risk Decision Support System) which uses a two-step scenario construction approach to handle the highly uncertain conditions of a disaster environment, coupled with an optimization model and robustness stress testing. We address the above challenges by exploring the potential of generative Artificial Intelligence (AI) to adapt the underlying methodology ad hoc in the disaster response phase to random disaster environments. This includes, for example, the reactive formulation of the optimization model according to the preferences of the decision makers and the application-specific development of scenarios that capture both the uncertainty and the dynamics of the disaster environment. The integration of AI is the next step in the development of ReDRiSS from an early prototype stage to a customized approach.

4 - Exploring Sustainable, Community-Focused Cooperative Ecosystems among Food Small-Medium Enterprises

Sofia Perez-Guzman, Georgia Insitute of Technology, Atlanta, GA, United States, Llord Brooks, Seckin Ozkul

The well-being of poor and underserved communities primarily depends on their direct access to vital resources in their vicinity. However, research concerning societal well-being and sustainability is often limited to the influence of large enterprises, while the role of small-medium enterprises (SMEs) still needs to be expanded. As such, this study aims to assess the generalization of inter-organizational systems designed to engender cooperative ecosystems that encourage inclusion, address bias towards SMEs, and impact how they serve their communities. Specifically, we focus on addressing how implementing food hubs and group purchasing systems among SME retailers can influence their competitiveness and the well-being of their communities. Thus, we develop a family of mathematical models for the logistics decision-making of the SMEs. We extend the profit maximization objective to include the impact of increased food access in the community. Furthermore, we explore conditions influencing the long-term sustainability of the cooperative ecosystems and the trade-offs between SME competitiveness and community well-being.

TD56

Summit - 448

Advancements in Digital Twin Technologies

Contributed Session

Chair: Jaeseok Yoon, N/A

1 - A Digital Twin of Outpatient and Day Care Services in Kuwait Cancer Control Center: Data to Action

Sulaiman Aman, Kuwait University, Kuwait, Kuwait, Abdullah Alibrahim, Abdulwahab Al-Tourah, Waddah Al-Refaie

Digital Twin for Health (DT4H) technology is an emerging tool that provides an empirical foundation for optimizing healthcare systems through simulated models of actual systems. By leveraging real-time data, dynamic analysis, and validated models, DT4H can profoundly enhance operational processes in cancer care by creating no-risk and cost-effective environment for conducting what-if analyses. This technology enables healthcare providers to revolutionize patient flows, allocate medical resources more efficiently, and improve scheduling procedures, thereby enhancing patient care. This presentation describes a model-driven approach using data integrations, observations, and models to construct a DT4H to improve cancer patient care, streamline processes, and optimize operations at the Outpatient Department (OPD) in the Kuwait Cancer Center. A data collection plan was implemented to capture patients' arrival rates, visiting frequencies, medical resource capacities, and administrative requirements. Continuous meetings with stakeholders established face validity and inspired an investigation of expanding afternoon clinics in response to increasing demand for cancer care, which is currently operating above its theoretical capacity. All collected data was integrated within a user-friendly discrete-event simulation (DES) model developed on Anylogic Simulation Software. The DES model incorporates animations and live visuals of performance indicators, such as waiting times and number of patients served. Findings from the digital twin indicated that increasing breast and gastrointestinal cancer afternoon clinics could decrease the average waiting time by 32% and increase the number of patients served daily by 8.6%. These results highlight DT4H's potential to elevate healthcare delivery through real-time adjustments and operational decision aids.

2 - Proactive Digital Twins with Virtual Sensors

Mark Mahyar Nejad, University of Delaware, Newark, DE, United States

Digital Twins are a virtual representation of real-world systems and can serve as a testbed for developing and evaluating optimal strategies in real time. A significant hurdle to their development and implementation is the volume of data required, and consequently the volume of sensors required. This paper proposes to reduce this volume through the creation of "Virtual Sensors" - machine learning models which use data from their geographic neighbors to impute the state of the network at their position, considering their relative position and the time to travel between them. These spatiotemporal considerations enable the creation of realistic, detailed, context-specific data, as opposed to traditional data generation. To create these models, we propose a spatiotemporal adaptation to the traditional Variational Autoencoder (VAE) means for data generation. In our model, the virtual sensor trains a VAE on its neighbor's data. Instead of sampling randomly from the latent distribution like the traditional VAE, our model learns the relation between its neighbor's current points in the latent space and draws from the space between them. In addition, we propose a novel physics-informed deep learning to supplement sensor data, seamlessly integrating fundamental principles of network dynamics to reduce bias and increase equity through representation in sensor sparse regions while reducing the massive data requirements of conventional deep learning strategies. By considering the geographic and temporal relations between its neighbors, the Spatiotemporally-Aware Variational AutoEncoder (STAVE) generates realistic, context-specific data, filling gaps in sensor networks.

3 - Evaluation of Simulation-Based Prediction Algorithm for Digital Twin Certification in the Urban Gas Industry

Jaeseok Yoon, Korea University, Seoul, Korea, Republic of, Sung Yeon Kim, Seonghyeon Han, Jinmin Kim

In the urban gas industry, price volatility, supply chain complexity, and enhanced environmental regulation are challenges facing this industry. This can be overcome by lowering the risk of the supply chain and increasing operational efficiency while responding to environmental regulations through certification. Therefore, in this study, we aim to evaluate the predictive algorithm for the gas industry digital twin certification. In addition, we compare the pressure fluctuation predictive performance using real-time data and simulation-based synthetic data to suggest its contribution to improving the operational efficiency, safety, and sustainable development of the gas industry. Our results highlight the excellent predictive performance of the autoregressive integrated moving average (ARIMA) model. Compared to raw data, it showed the best performance in all evaluation indices, mean absolute percentage error (MAPE), root mean square error (RMSE) and determinant (R2). These results demonstrate that the integration of digital twin technology and predictive models can revolutionize the maintenance strategy, operational efficiency, and risk prediction of the gas industry. Predictive maintenance models can help prevent significant industrial risks such as gas leakage accidents. Furthermore, the innovative method of implementing the proposed digital twin technology and the proposed predictive model lay the theoretical basis for certification of the application of digital twins in the gas industry in the future.

TD57

Summit - Terrace Suite 1

Revolutionizing Healthcare: Innovations in Analytics and Operations

Invited Session

Health Applications Society

Chair: Liang (Leon) Xu, Singapore Management University, Singapore, 178899, Singapore

1 - Inducing Compliance with Good Machine Learning Practices In Developing Adaptive AI-Enabled Medical Devices

Liang (Leon) Xu, Singapore Management University, Singapore, Singapore, Jiayi Lai, Xin Fang, Tinglong Dai

This paper presents a two-period regulatory framework for the development and market clearance of adaptive AI-enabled software as a medical device (SaMD). The framework explores the interactions between the developer and the regulator, focusing on the developer's adherence to Good Machine Learning Practices (GMLP) and the regulator's role in ensuring compliance.

In the first period, the developer trains an AI/ML algorithm, and the regulator bases the initial approval decision on the reported efficacy and whether the developer adheres to GMLP. In the second period, the developer can retrain the algorithm using real-world data to potentially improve efficacy. The regulator then assesses compliance during retraining before granting re-clearance.

Using a Perfect Bayesian Equilibrium model, this study analyzes the decision-making dynamics between the developer and the regulator under different regulatory paths, namely, adaptive algorithm with re-inspection, pre-determined change plan, and locked algorithm. The insights provided aim to balance innovation with compliance, ensuring responsible development and deployment of AI-enabled SaMD while safeguarding patient safety.

2 - From Black to Grey: Improving Access to Antimalarial Drugs in the Presence of Counterfeits

Jiatao Ding, INSEAD, Singapore, Singapore, Michael Freeman, Sasa Zorc

In malaria-endemic countries, the limited availability of affordable antimalarial medication has contributed to the widespread distribution of inferior counterfeit drugs. We study such markets to determine how philanthropic donors can best allocate limited funds to subsidize the purchase or sales of antimalarial drugs via private-sector distribution channels. We also consider the potential effectiveness of non-subsidy interventions, e.g., traceability technology adoption, to improve outcomes in the presence of counterfeit drugs. To examine the supply chain of antimalarials, we develop a game-theoretic model in which the retailer has a strategic choice to source legitimate drugs from a certified supplier, potentially counterfeit drugs from an uncertified supplier, or both. In contrast with the extant literature, we show that in the presence of counterfeits, employing a purchase subsidy alone may no longer be optimal. In particular, when the donor's budget is small, the donor may prefer to offer a sales subsidy that covers both legitimate and counterfeit drugs; if the drug's retail price is exogenous and demand uncertainty is high, the donor may need to refrain from offering any subsidy at all. Moreover, we identify conditions under which traceability technologies may or may not help improve the outcomes. Finally, we perform an extensive numerical analysis, calibrating the models to malaria data from Mozambique. Our results indicate the need for regulators and donors to understand specific market characteristics in order to design effective subsidy schemes and select appropriate technologies and policies to improve access to life-saving medicines.

3 - Sharing in Caring: Analyzing Shared Medical Appointments Through An Operations Lens

Sundara Natarajan Panchanatham, HKU Business School, Hong Kong, Hong Kong, Enver Yucesan

Shared medical appointments (SMA) are visits in which healthcare professionals see groups of patients in a concurrent session. The setup can enhance patient experience, eliminate redundant work, and reduce costs for specific treatments. We mathematically model the operational trade-offs in SMAs and derive optimal capacity decisions using a game-theoretic framework.

4 - Medical Inequality: Helping Vulnerable Hospitals

Qiang Li, Wilfrid Laurier University, Waterloo, ON, Canada, Wei Gu, Meng Li

Healthcare resources in developing countries are often unevenly distributed, posting challenges for hospitals to provide essential medical services and supplies equitably to their communities. In response, governments and nonprofit organizations have implemented various programs to support underresourced hospitals located in remote areas. We leverage the introduction of external experts to an underresourced hospital and employ the difference-in-differences design to assess the program's impact. We find that the program has significantly improved the efficiency and quality of medical services provided by them.

TD58

Summit - Terrace Suite 2

Pengyi's session

Invited Session

Health Applications Society

Chair: Pengyi Shi, Purdue University, West Lafayette, IN, 47907, United States

1 - Safe Reinforcement Learning with Contextual Information: Theory and Application to Personalized Comorbidity Management

Esmaeil Keyvanshokooh, Mays Business School, Texas A&M University, College Station, TX, United States, Junyu Cao, Tian Liu

Optimizing the treatment regimen is a crucial sequential medical decision-making problem. Inspired by this problem, we study a sequential decision-making setting of learning a personalized, safe control policy that maximizes an objective function subject to safety constraints that need to be satisfied during the learning process. We develop a model-based reinforcement learning (RL) algorithm that accounts for (i) personalization, (ii) safe exploration, and (iii) general statistical models for modeling uncertainty, with a rigorous regret performance guarantee. We utilize a granular clinical dataset of patients with co-morbid type 2 diabetes and hypertension to evaluate the performance of our theory.

2 - Optimizing HIV Prevention Funds Allocation to Minimize New HIV Cases – It does matter!

Md Hafizul Islam, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States, Evin Jacobson, Ram Shrestha, Paul Farnham

Local U.S. health jurisdictions receive funds from the Centers for Disease Control and Prevention (CDC) to implement a comprehensive HIV prevention plan to reduce HIV incidence and improve health equity across their community. We developed a linear programming model combined with a Bernoulli process model to identify the optimal allocation of the CDC's HIV prevention funding for local health jurisdictions. The model allocates the CDC provided budget across different HIV prevention interventions serving different populations stratified by gender and transmission groups to maximize the number of new HIV cases prevented. Inputs to the model include available budget, the maximum number of persons that could be reached by the HIV prevention intervention (reach limits), and data on the costs and efficacy of HIV prevention programs for different subpopulations. In a case study, we applied the model using data from a specific health jurisdiction and analyzed the resulting solutions. We solved the model for a scenario where the model is restricted to follow the allocation practice reported to the CDC by the health department and an unrestricted scenario where the model is allowed to allocate funds to any prevention intervention within the reach limits. Comparing the optimal funding allocations recommended by the model for both scenarios suggests that there is a potential for improving the number of new HIV cases averted when considering alternate funding allocations.

3 - Pooling in-Person and Virtual Queues with An Application to Telehealth

Omer Berk Olmez, Baruch College, Graduate Center, CUNY, New York, NY, United States, Alex Mills

Healthcare services can be offered both in-person and virtually by the same providers. We study a clinic's decision of whether to pool the resources of both channels or to have dedicated resources for each channel. We consider a model in which strategic patients actively decide which channel to join and balk while observing the waiting time for both channels and the distance to the in-person facility. Contrary to the consensus in the classical pooling literature, our findings indicate that patients' expected waiting times and net utilities do not necessarily improve with pooling. Furthermore, our research highlights that pooling the two channels yields benefits for the providers only when the net profit generated from the virtual channel is substantial.

TD59

Summit - Ballroom 1

Energy-Aware Business Practices

Contributed Session

Chair: Seongkon Lee, Korea Institute of Energy Research, Daejeon, Korea, Republic of

1 - Market and Corporate Status Analysis for a self-sustaining Smart Thermal Energy Design Platform

Seongkon Lee, Korea Institute of Energy Research, Daejeon, Korea, Republic of, Mingye Lee, Wookhyun Lee, Sangkon Lee, Daekeun Lee, Gento Mogi

The importance of energy technology is becoming more and more important for carbon neutrality and addressing climate change issues. Among energy technologies, especially in the industrial field, improving energy efficiency is one of the major issues. We plan to build a smart Thermal Energy Design Platform(STED) to strengthen the product design and production capabilities of small and medium-sized companies in the energy-consuming manufacturing industry. In order to build a self-sustaining STED, this study focuses on the analysis of market and corporate status in thermal energy-intensive industry area from the industrial ecosystem perspective and draws implications for self-sustaining operation of the STED.

Key words Market analysis, Corportate Status Analysis, Thermal Energy-intensive Industry, Value Stream Analysis

Acknowledgements This research is carried out by the support of Ministry of Trade, Industry, and Energy(MOTIE) and KETEP (No. 20202020800200), Republic of Korea.

2 - A Note on ESG CAPM, Factor Models, and Carbon Reduction Index Funds

Hiroshi Ishijima, Chuo University, Tokyo, Japan, Akira Maeda

Global efforts to achieve net-zero carbon emissions by 2050 and mitigate climate change are intensively driven by capital market functions, including ESG investing. In this context, we develop an asset pricing model for ESG investing that extends the conventional model (Cochrane 2005). In our model, the asset price is the sum of the present value of future cash flows and ESG dividends. ESG dividends are a bundle of benefits related to environmental, social, and governance factors, such as reducing carbon emissions. Because ESG dividends are non-financial, they are converted to monetary value using the extended stochastic discount factor. Our model represents a concept called the double bottom line for ESG investing. We also derive several versions of the model, namely ESG consumption CAPM, ESG CAPM and ESG factor model. In addition, we show the "Green-Brown-Mix Irreverent Proposition" for ESG investing. It asserts that the sum of risk-adjusted excess returns (brown) and ESG dividends (green) for each asset must be identical to that of the market benchmark portfolio. In the literature, much empirical work - including Bolton and Kacperczyk (2021) - has been devoted to finding a double bottom line in the market. This study serves as a foundation for empirical research to show whether or not ESG investing allows us to achieve a double bottom line. As a typical application, we create the Carbon Reduction Index Fund, which invests in Japanese companies committed to reducing carbon emissions. We show that the fund delivers superior investment performance.

3 - Carbon Tracker Features and Accuracy to Measure Carbon Emissions of Algorithms

Eric Van Heck, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Max Baiwir, Anna Priante

Many algorithms are developed in the information systems community. Algorithms are helping but also harming the Earth. They can help by creating business models to reduce waste, emissions, and pollution, but also can harm by emitting carbon emissions that will cause global warming. The use of carbon trackers, i.e. the monitoring and measurement tools, can minimize carbon footprints. This study aims to evaluate carbon tracker features on the accuracy of carbon measurements for algorithms. Three important features are investigated: real-time monitoring of the energy consumption, real-time monitoring of the carbon intensity of its energy sources, and the level of monitoring automation and integration. In a benchmark study, we compare three carbon trackers in terms of the accuracy to measure carbon footprints: Green Algorithms, CodeCarbon, and Tracarbon. These three carbon trackers were evaluated using the Residual Network (ResNet) algorithm with increasing levels of complexity and the MNIST data set. Tracarbon is most accurate for the lower complex and higher complex algorithms. For the ResNet-44 algorithm CodeCarbon is very accurate and for the ResNet-56 algorithm Green Algorithms is reasonably accurate. Green Algorithms stands out for its superior accuracy in estimating the carbon intensity value and underestimates the energy consumption. Conversely, CodeCarbon tends to overestimate energy consumption but is on average better in measuring carbon intensity. Tracarbon provides precise estimates for energy consumption and underestimates the carbon intensity. Recommendations of carbon tracker usage and carbon footprint reduction strategies for the members of the INFORMS community will be presented.

4 - Technology-driven decision-making forecast to enhance sustainability in the Iron and Steel Industry

Hanseul Jo, Kyung Hee University, Yongin, Korea, Republic of, Jungwoo Shin

With the growing importance of carbon neutrality, industries are undergoing significant changes. The transition to carbon-zero industries imposes substantial technological and financial burdens, requiring changes in raw materials, production process, and technologies. Consequently, the decision-making process for selecting new carbon-free technologies has become a critical issue for the industrial sectors. Governments also grapple with integrating such industry-level decision-making process into national carbon neutral strategies. By considering the industrial levels, governments can assess the appropriateness of their objectives and accelerate carbon neutrality by conducting more relevant supportive policies.

Based on this importance, this study predicts how global technology transition occurs in the iron and steel industry by combining the revealed preference (RP) data and stated preference (SP) data. In particular, the iron and steel industry is highly important as it is one of the largest carbon producing industries. To estimate the technology diffusion trajectories and learning effect of each emerging technology and country, RP data is used to figure out the current technological levels and states of national carbon neutralization. SP data is used to examine experts' forecasts on the global technology transitions and differences in technology development of each country.

By providing a comprehensive approach to the industry-level decision-making process and incorporating them into national roadmaps, this study provides meaningful insights into upcoming carbon-zero technology transitions. These make stakeholders conduct strategic decision-making and investment for carbon neutrality. Ultimately, both stakeholders and governments can gain valuable insights into potential pathways for achieving sustainability goals.

5 - "Unleashing Green Logistics: Key Factors Driving Sustainability in South Korean Logistics SMEs"

seung hee ha, Kyungpook National University, Daegu , Korea, Republic of, Sungsu Kim

Driven by stringent carbon emission regulations and environmental conservation efforts globally, enterprises across sectors are actively embracing sustainable management practices. While large corporations have spearheaded this trend, small and medium-sized enterprises (SMEs) are also experiencing mounting pressure to conform due to stakeholder influences and network relationships. The logistics industry, with its extensive operational scope from raw material procurement to product delivery, faces heightened emphasis on environmentally friendly practices and regulations. However, SMEs in this sector grapple with significant challenges in implementing green practices, stemming from technological constraints, financial limitations, and concerns over business continuity. Research specifically targeting SMEs in the logistics industry remains limited.

This study aims to identify and prioritize the key determinants influencing the adoption of green practices among Korean logistics SMEs. The variables under investigation are categorized into external factors, grounded in stakeholder theory, such as government regulations, supplier cooperation, and community responsibility, as well as an internal factor: the company's green technological readiness. These variables are hypothesized to influence the level of green logistics activities through the mediating variable of introduction intention. Furthermore, the study evaluates the performance implications of green logistics adoption levels.

Through this comprehensive analysis, the study seeks to contribute to practical policy formulation by prioritizing the determinants, thereby enhancing the perception of SMEs hesitant to embrace green management strategies. Additionally, it endeavors to verify the contribution of green practices to business survival and sustainable management practices within the Korean logistics industry.

Keywords: Green Logistics;Korean Logistics SMEs; Stakeholder Theory;Green Practices Adoption;Environmental Sustainability

TD60

Summit - Ballroom 2

Inventory placement and Inbound Network Design

Invited Session

The Practice Section of INFORMS

Chair: Chun Ye, Amazon.com, Seattle, WA, United States

Co-Chair: Zihao Li, Amazon, Bellevue, WA, United States

1 - Probabilistic Approach to Large-Scale Inventory Placement

Cristiana Lara, Amazon, Bellevue, WA, United States, Arash Haddadan, David Mildebrath, R Ravi

Amazon offers millions of unique items for sale on its website. The popularity of these items is not evenly distributed, and the mean weekly demand for different products on Amazon spans multiple orders of magnitude. Because of this enormous scale and selection, and asymmetries in our storage nodes, deciding how to optimally place this inventory throughout our fulfillment network is very challenging. Directly modeling the massive number of products into a monolithic formulation is intractable. On the other hand, downsampling products typically does not adequately capture the behavior of the “long tail”. We propose a scalable method for grouping products into groups based on their popularity. As we aggregate distinct products into a single group, we lose information about their specific demand and inventory, which results in overly optimistic plans. To mitigate this issue, we propose a “Selection Spread” function, a probabilistic approach to model non-fungibility of products within a group. Our model matches the shipping and storage capacity of nodes in our network with the velocity profile of demand in order to minimize fulfillment costs. We use this framework to derive insights on the relationship between placement strategy and customer experience.

2 - Evaluating Regional Decentralization of Inventory in Delivery Speed-Sensitive Retail Networks: A Comparative Study between Newsvendor Approximation and Simulation

Katja Meuche, Georgia Institute of Technology, Atlanta, GA, United States, Yuan Li, Amitabh Sinha, Benoit Montreuil

Large retailers such as Amazon offer through online channels a vast array of products, encompassing millions of unique items. The multitude of products creates a complex demand landscape where high-demand products are categorized as "head," moderate-demand as "body," and low-demand as "tail." To control complexity and achieve efficiency in the fulfillment network, Amazon has recently regionalized its operations in the US. This process involved defining the number and boundaries of regions, assigning fulfillment centers (FCs) to regions, and aiming to fulfill customer demands using the inventory of the FCs within each region versus from FCs in nearby regions or centralized FCs. This prompts a crucial question: What level of regional decentralization of inventory is economically beneficial for each product category when considering the customer's sensitivity to delivery speed?

In this talk, we analyze this question in two ways for the continental US with alternative inventory de/centralization in one to 48 regions. First, we present a Newsvendor model that considers transportation costs and order-to-delivery-time sensitivity. Our results show that more decentralization of head and body products tends to achieve more profit maximization, and the inverse for tail products. Then, we compare the results of the Newsvendor model to results from a simulation that dynamically models fulfillment of online channel demand in more detail. Based on the comparison, we determine how well the adapted Newsvendor model approximates online retail operations. We then draw insights and conclusions relative to regional de/centralization of inventory considering time-sensitive customers.

3 - A Unified Bulk Storage and Distribution Network Design

Shanshan Zhang, Amazon, Bellevue, WA, United States

Amazon operates several types of bulk distribution centers, each tailored to specific operational models and network designs. Despite their individual nuances, their major purpose remains consistent: to provide low-cost storage solutions by handling inventory in bulk formats while facilitating fast replenishment to their downstream premium storage nodes positioned proximally to Amazon end customers. Recognizing the inherent synergies within these operations, network designs, and technological requirements, this presentation presents a unified bulk storage and distribution network design strategy to streamline operations and enhance efficiency across the network.

4 - Ideal Inventory Placement with Demand Spillover

Zihao Li, Amazon, Bellevue, WA, United States, Zhongxiang Wang, Tolga Cezik

We share an ideal inventory placement strategy for designing an effective inbound network of an online retailer. In an online retailer's fulfillment network, demand is ideally fulfilled from the nearest warehouse with inventory and ship capacity before other, further-away warehouses (defined as spillover). The spillover of demand leads to variable (inventory) turn (measured as the ratio of the average daily shipment units to the average on-hand units) for each item at different warehouses. We build a two-stage stochastic model to capture the variable turns of an item at different warehouses in inventory placement and a discrete-event simulation model to assess its impact on fulfillment. We generate two inventory allocation plans under two inventory placement strategies: assuming all warehouses have the same network turn (Constant Turn) for an item or incorporating the variable turns (Variable Turn) at warehouses for an item. Simulation is used to evaluate the fulfillment of the two inventory placement strategies, showing that Variable Turn reduces total fulfillment distance.

TD61

Summit - Ballroom 3

Professors and Their Startups: Reflection on Academic Entrepreneurship

Panel Session

Committee’s Choice

Co-Chair: Serguei Netessine, The Wharton School, Philadelphia, PA, United States

1 - Panelist

Srikanth Jagabathula, NYU Stern School of Business, New York, NY, United States

2 - Panelist

Gürhan Kok, Koc University, Istanbul, Turkey

TD62

Summit - Signature Room

Digital Transformation in Transportation Systems: Navigating User Behavior and System Efficiency

Invited Session

TutORial

Chair: Jamol Pender, Cornell University, Ithaca, NY, United States

1 - Digital Transformation in Transportation Systems: Navigating User Behavior and System Efficiency

Haripriya Pulyassary, Cornell University, Ithaca, NY, United States, Manxi Wu

This tutorial explores the impact of digitalization on today’s transportation systems, focusing on how emerging information technologies and pricing schemes are reshaping the travel behavior in congestible networks. Our focuses include: static and dynamic routing games, the impact of asymmetric information on network efficiency, and the design of incentives for system efficiency and equity. Through a combination of theoretical insights and empirical studies, the tutorial offers an in-depth analysis of models and tools for analyzing strategic user behavior, the role of information and incentive mechanisms in promoting socially desirable outcomes, and the application of these theories in real-world transportation systems.

TD63

Regency - 601

Platform Design in the Digital Era

Invited Session

Information Systems

Chair: Xiangjing Chen, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - The Tale of Two Channels: Monetizing Free and Members-Only Content on a Video-Sharing Platform

Dongsheng Li, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Yi-Jen Ian Ho, Xiangjing Chen, Shengjun Mao

Video-sharing platforms, such as YouTube, enable content creators to monetize their videos through both advertising commissions on free channels and subscription fees on paid channels. Despite the prevalence of this monetization model, the strategic interactions among creators, consumers, and a platform remain unclear. To address this gap, we develop a game-theoretical model to study how creators optimally allocate their efforts towards content quality and set subscription fees to compete in a duopoly setting. Our analysis provides clear guidance for content creators and valuable insights for consumers.

2 - Political Impact of Social Media Fake Accounts

Zihong Huang, Texas Tech University - Rawls College of Business, Lubbock, TX, United States

Fake accounts are prevalent on social media platforms to manipulate people's opinions. Many anecdotal evidences suggest that fake accounts can impact the outcome of political campaigns. However, there is a lack of understanding of how political manipulation by fake accounts works. In this study, we incorporate two competing parties, the potential voters, and the social media platform into a game theoretical model to yield the parties' equilibrium competing strategies. In addition, we also examine the impact of the platform's anti-fake strategy on the political outcome and social welfare.

3 - Regurgitative Training: the Value of Authentic Data in Training Generative Ai Models

Jinghui Zhang, Tsinghua University, Beijing, China, People's Republic of, Dandan Qiao, Mochen Yang

What happens if we train a new Generative AI (GenAI) model using data that are at least partially generated by other GenAI models? The explosive success of GenAI models, such as ChatGPT means that a substantial amount of content online will be generated by AI rather than humans, which will inevitably enter the training datasets of next-generation GenAI models. In this paper, we evaluate the implications of such “regurgitative training” on GenAI performance. Through fine-tuning GPT-3.5 with data generated by itself in a machine translation task, we find strong evidence that regurgitative training clearly handicaps the performance of GenAI models. The ease of getting large quantities of AI-generated data does not seem to compensate for performance loss — even training with a fraction of authentic data is enough to outperform regurgitative training. Furthermore, we also explore several different mitigation strategies to reduce the performance losses of regurgitative training. These strategies can improve the performance of regurgitative training to some extent but are not able to fully close the gap. Overall, our work highlights the value of authentic, human-generated data in training GenAI models, which cannot be easily substituted by AI-generated data.

4 - The Fairy&Rsquo;S Magic for Pinocchio&Rsquo;S Nose: Designing Return Policies Against Review Manipulation

Xiangjing Chen, University of Nebraska-Lincoln, Lincoln, NE, United States, Yi-Jen Ian Ho, Shengjun Mao

Review manipulation is pervasive across online platforms, despite the existence of various technologies aimed at detecting fake reviews. Review manipulation remains rampant as platforms may lack economic incentives to mitigate the problem. While previous literature focuses on developing algorithmic solutions, we take a different approach by examining the economic incentives of platforms through their return policies. We craft a game-theoretic model, endogenizing a platform's return policies and sellers' review manipulation efforts and pricing, given heterogeneous consumers in preferences. Our results reveal that the impact of a full-refund policy on review manipulation varies depending on the level of competition among sellers. We further uncover a return-manipulation paradox: a platform is more inclined to choose the policy that encourages review manipulation. To address this paradox, we propose an autonomous scheme that can effectively lower overall manipulation level and increase social welfare.

TD64

Regency - 602

Social Media and Digital Platforms

Invited Session

Social Media Analytics

Chair: Wenqi Shen, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Alice Jang, Virginia Tech, Blacksburg, VA, 24061, United States

1 - Managing the Transition from Crowd Volunteerism to Paid Work: Evidence from the Field

Jiho Lee, Boston University, Boston, MA, United States, Gordon Burtch, Chrysanthos Dellarocas

Many digital platforms, particularly at their outset, rely on volunteers' contributions to create products and services of economic value. Often, these organizations subsequently elect to transition toward a system of paid work, for a variety of reasons, e.g., ethical, legal, or logistical considerations, leading to questions of how platforms should navigate the transition. Of particular importance are human capital considerations, given differences in the motivations, perspectives, and behavior of volunteers versus paid workers. To understand the determinants of volunteers' interest and motivation (or lack thereof) to engage as formal employees, we leverage data from a large online platform for language education, which initially relied on volunteers for course and lesson development before transitioning to a closed model of paid work. We first examine the factors that determine volunteers' interest in pursuing paid positions, drawing on prior theories of volunteer motivations. Subsequently, we examine how the same set of factors relates to the performance of volunteers who are eventually hired. We find that a volunteer's work patterns, their possession of a leadership role, and social considerations (i.e., their peers' interest in remaining with the organization) are all important determinants of a volunteer's desire to pursue a paid position. At the same time, we find that these same factors have varied effects on a volunteer's eventual performance in paid work. We corroborate our quantitative results using employer-conducted surveys of the paid workers, and we discuss implications for platform operators.

2 - Group sentiment polarization of social media users during emergent events: insights from China

Bairong Wang, Shanghai Maritime University, Shanghai, China, People's Republic of

This study analyzes the dynamics of social media users' sentiments by conducting 8 case studies of emergent events in China. Over 20 thousand Sina Weibos were analyzed to explore the features of users' comments, sentiment, and information-spreading behaviors, as well as how official media and grassroots media behave in leading the development of group sentiment. The results indicate that small groups present severe polarization of their sentiment and opinions, and within their discussions, limited contradicting opinions are presented. Compared to the official media accounts, grassroots media accounts are more influential on users' sentiments and opinions. However, bot-susceptible users are found to spread polarized sentiment comments, which may accelerate the polarization of group sentiment and opinions. Managerial insights are discussed.

3 - Emotional Contagion and Attenuation in Live Streaming Platform

Alice Jang, Virginia Tech, Blacksburg, VA, United States

Live streaming platforms, like Twitch, offer a space for real-time interaction among users, forming the basis for online communities. Although emotional contagion has been studied in asynchronous social media platforms, its dynamics in synchronous environments such as live streaming platforms remain less understood. These platforms feature synchronous communication and extensive use of non-textual emojis. By analyzing a large dataset of chatroom interactions on Twitch, specifically focusing on emoji usage, this study explores how emotional contagion and attenuation influence user engagement within these live streaming communities.

TD65

Regency - 603

Data-driven Decision-making: Personalization, Practicality, Perspicuity

Invited Session

Artificial Intelligence

Chair: Linda Zhao, University of Pennsylvania, Philadelphia, PA, United States

1 - Personalized reinforcement learning: with applications to sepsis management in ICU

Junhui Cai, University of Notre Dame, Notre Dame, IN, United States, Ran Chen, Martin Wainwright, Linda Zhao

Reinforcement learning (RL) has achieved remarkable success across various domains; however, its applicability is often hampered by challenges in practicality and interpretability. Many real-world applications, such as in healthcare, have large and/or continuous state and action spaces and demand personalized solutions. In addition, the interpretability of the model is crucial to decision-makers so as to guide their decision-making process while incorporating their domain knowledge. To bridge this gap, we propose a personalized reinforcement learning framework that integrates personalized information into the state-transition and reward-generating mechanisms. We develop an online RL algorithm for our framework. Specifically, our algorithm learns the embeddings of the personalized state-transition distribution in a Reproducing Kernel Hilbert Space (RKHS) by balancing the exploitation-exploration tradeoff. We further provide the regret bound of the algorithm and demonstrate its effectiveness in sepsis management in ICU.

2 - Doubly high-dimensional contextual bandit: An Interpretable Model with Applications to Assortment/Pricing

Ran Chen, Washington University in St. Louis, St. Louis, MO, United States, Junhui Cai, Martin Wainwright, Linda Zhao

Key challenges in running a retail business include how to select products to present to consumers (the assortment problem), and how to price products (the pricing problem) to maximize revenue or profit. Instead of considering these problems in isolation, we propose a joint approach to assortment-pricing based on contextual bandits. Our model is doubly high-dimensional, in that both context vectors and actions are allowed to take values in high-dimensional spaces. In order to circumvent the curse of dimensionality, we propose a simple yet flexible model that captures the interactions between covariates and actions via a (near) low-rank representation matrix. The resulting class of models is reasonably expressive while remaining interpretable through latent factors, and includes various structured linear bandit and pricing models as particular cases. We propose a computationally tractable procedure that combines an exploration/exploitation protocol with an efficient low-rank matrix estimator, and we prove bounds on its regret. Simulation results show that this method has lower regret than state-of-the-art methods applied to various standard bandit and pricing models. Real-world case studies on the assortment-pricing problem, from an industry-leading instant noodles company to an emerging beauty start-up, underscore the gains achievable using our method. In each case, we show at least three-fold gains in revenue or profit by our bandit method, as well as the interpretability of the latent factor models that are learned.

3 - Transfer Learning with Random Coefficient Ridge Regression

Hongzhe Zhang, University of Pennsylvania, Philadelphia, PA, United States, Hongzhe Li

Ridge regression with random coefficients provides an important alternative to fixed coefficients regression in high dimensional setting when the effects are expected to be small but not zeros. This paper considers estimation and prediction of random coefficient ridge regression in the setting of transfer learning, where in addition to observations from the target model, source samples from different but possibly related regression models are available. The informativeness of the source model to the target model can be quantified by the correlation between the regression coefficients. We propose two estimators of regression coefficients of the target model as the weighted sum of the ridge estimates of both target and source models, where the weights can be determined by minimizing the limiting estimation risk or prediction risk. Using random matrix theory, the limiting values of the optimal weights are derived under the setting when p/n goes to gamma, where p is the number of the predictors and n is the sample size, which lead to an explicit expression of the estimation or prediction risks. We will show that the transfer learning prediction based on estimation risk is robust to distributional shift of the predictors. Simulations show that these limiting risks agree very well with the empirical risks. An application to predicting the polygenic risk scores for lipid traits shows such transfer learning methods lead to smaller prediction errors than the single sample ridge regression or Lasso-based transfer learning.

TD66

Regency - 604

AI Application in Education

Invited Session

Artificial Intelligence

Chair: Zhe Deng, Saint Joseph's University, 5400 City Ave, Media, PA, 19131, United States

1 - Multimodal Video Learning: Managing Cognitive Load Patterns using Guided Attention Graphs

Yi-Yun Chu, Carnegie Mellon University, Pittsburgh, PA, United States, Zhe Deng, Wen Wang, Pedro Ferreira

This study explores the intricacies of cognitive load management within video-based learning environments, which have become pivotal in education through their integration in massive open online courses (MOOCs) and asynchronous learning models. The proliferation of technology-enhanced learning (TEL) environments accommodates diverse learning preferences globally, integrating auditory, visual, and textual elements. This requires efficient cognitive load management, which involves balancing intrinsic load (related to content complexity), extraneous load (influenced by presentation style), and germane load (associated with learner engagement). Challenges in managing cognitive load are magnified by design elements such as visual and audio quality, which may inadvertently elevate cognitive strain. To mitigate information overload, online platforms have adopted adaptive learning systems, summarization tools, and attention-guiding graphs. This study aims to construct a novel methodology for measuring cognitive load in video-based learning environments and to explore how the interplay between cognitive load and attention graphs affects learning outcomes. By enhancing cognitive load management and improving learner interaction with educational content, this research seeks to boost the effectiveness of video-based learning.

2 - AI in Education and Machines as Thought Partners: an AI researcher's perspective

Marcello Balduccini, Saint Joseph's University, Wynnewood, PA, United States

In this presentation, we will look at AI in education from the perspective of an AI researcher involved in developing cutting-edge technologies but who also needs to confront, as a teacher, the reality of the impact of AI in education. We will begin by talking about the various areas of AI and about the relation each of them has with education. Among those, we will necessarily talk about Large Language Models and their potential in education, but also all the issues they raise. We will conclude the talk by discussing how the latest advancements in AI usher a possible era of "machines as thought partners" and how AI breakthroughs that are possibly just below the horizon may affect education in the near future.

3 - Generative AI-Powered Educational Alignment: A Framework for Matching Syllabus Course Topics with Web Description

Ruben Mendoza, Saint Joseph's University, Philadelphia, PA, United States

With the burgeoning interest and advancements in generative artificial intelligence (AI), its application in the educational sector is increasingly gaining attention from researchers. This study explores a multifaceted dimension of higher education, extending beyond mere students' course registration to encompass instructors' course preparation and the broader implications for university branding: the congruence between online course descriptions and the actual course syllabi. However, discrepancies between the course catalog descriptions and the actual course content, as detailed in the syllabus, can lead to reduced learning efficiency, student disillusionment, misaligned academic expectations, and challenges in academic planning and progress. To tackle this issue, our research introduces a novel framework utilizing generative AI to systematically evaluate and rectify mismatches between course web descriptions and syllabus content. The methodology involves employing OpenAI's techniques to extract key topics from course syllabi. Subsequently, we utilize advanced embedding methods, including BERT, GPT-2, RoBERTa, and DistilBERT, coupled with cosine similarity metrics, to assess the congruence between the extracted topics and the course description content. The framework also integrates an outlier detection algorithm to identify courses with significant misalignments. Moreover, this study proposes the use of generative AI applications to refine and enhance course catalog descriptions. This approach not only aids higher education institutions in updating their course offerings with cutting-edge technology but also contributes to curriculum development. It fosters an environment conducive to improved student learning efficiency and more effectively designed courses, thereby enhancing the overall educational experience.

4 - Can Student Faces Tell Learning?

Zhe Deng, Saint Joseph's University, Media, PA, United States

We investigate the feasibility of predicting learning outcomes from the facial expressions of students while they engage with instructional videos or participate in remote sessions. The potential application of real-time facial expression analysis could enhance the personalization of educational sessions. Data for this study comes from several executive education sessions held in Lisbon, Portugal.

TD67

Regency - 605

AI & Robotics: Impact on Perceptions and Innovation

Contributed Session

Chair: Yixuan Zhang, University of Science and Technology Beijing, Beijing, N/A

1 - Effects of AI-Powered Review Responses on Consumer Perceptions of Service

Yixuan Zhang, University of Science and Technology Beijing, Beijing, China, People's Republic of, Yang Jiang, Xiangbin Yan

Artificial intelligence (AI) has increasingly become a tool for managing user interactions, yet there is limited understanding of its involvement in review responses. In this paper, we investigate the effect of utilizing AI technologies in managerial review responses on consumers' perceptions of service. Conducting experiments on an online platform, we obtain several interesting findings. We find that AI involvement in managing review responses positively affects users' perceived service quality, perceived service effort, and satisfaction. This impact differs when responding to positive reviews and negative reviews. Additionally, we consider the moderating factors of prompt types and the disclosure of AI involvement. Results show that providing AI with more rational cues is beneficial for negative reviews by enhancing the problem-solving ability of responses, whereas emotional prompts are suitable for responding to positive reviews to bolster the empathetic tone. However, the disclosure of AI involvement in review responses tends to reduce its positive impact on user perceptions and attitudes toward service. Our study provides valuable insights for understanding the role of AI in service management and has important managerial implications for firms seeking to deploy AI to improve service delivery and shape the future practice of management.

2 - High-Risk AI System for Business? Investigating the Passive Pathway of Stereotype Threat in AI-Human Interaction

Minki Kim, KAIST College of Business, Seoul, Korea, Republic of, Seonghun Yun, Dain Jung, Dowon Kwak

AI-induced harm to individuals and society calls for the ethical employment of AI and initiatives across the globe. However, scholarly discussions have mostly focused on the harm arising from AI and its algorithmic bias in active pathways. This study centers on biases in AI-human interactions, which stem not from the design of AI, but from users' misconceptions of AI agents with human-like features. For empirical identification, we exploit a unique experimental transition of educational tutoring programs in STEM and non-STEM domains from a human-to-tablet interface to a human-to-AI interface. Our research reveals that female consumers experience gender stereotype threats when interacting with male-characterized AI in the STEM domain, leading to negatively impacting the AI's benefits on their outcomes. Our robust results underscore the importance of risk management for AI systems to prevent accidental harm arising from consumers' self-perceived gender bias. Aligned with the recent emphasis on diversity, equity, and inclusion in the application of AI in the business domain, our findings provide managerial implications regarding potential passive negative effects when implementing AI services.

3 - Reviewing Forecast Results at Scale: An ML/AI-Driven Approach for Smarter Decisions

Sriram Sambasivam, Estee Lauder, Edison, NJ, United States

In the global Consumer Packaged Goods (CPG) industry, accurate demand forecasting is paramount for optimizing production, reducing waste, and meeting customer demands across diverse markets. This work focuses on reviewing forecast results generated by a machine learning/artificial intelligence (ML/AI) system for a typical CPG manufacturer with a worldwide presence, encompassing multiple product categories, channels, and brands.

The ML/AI system generates forecasts that are critical for strategic decision-making. However, these forecasts can sometimes contain anomalies, outliers, runaways, or misalignments in run rates, which can lead to suboptimal decisions if not identified and addressed promptly.

This work presents an innovative approach to reviewing forecast results at scale, leveraging advanced analytics, ML and AI techniques. By analyzing forecast data for anomalies and outliers, identifying potential runaways or misalignments, and applying statistical tests and machine learning models, this approach aims to enhance the accuracy, stability and reliability of forecasts.

To facilitate the implementation of this approach, we introduce an accompanying R/Python package. This package provides a framework for automated forecast review, allowing CPG manufacturers to efficiently and effectively identify and address issues in their forecast results.

TD68

Regency - 606

Advancements in Statistical Methods and AI

Contributed Session

Chair: Angela Morrison, University of Colorado - Denver, Denver, CO, United States

1 - One-at-a-time Knockoffs: Controlled False Discovery Rate with Higher Power

Charlie Guan, Northwestern University, Evanston, IL, United States, Zhimei Ren, Daniel Apley

Linear regression is a versatile tool ubiquitous in scientific analysis and machine learning. Consider the standard model y=Xβ+z where y is the response, X is the design matrix, β is the vector of unknown coefficients, and z is noise. We study the case where β is sparse, i.e., a small subset of β is nonzero. One application of this problem is genome-wide association studies where researchers seek to discover a small set of genes that affect cancer drug resistance from hundreds of thousands of genes. The objective is variable selection: recover the variables with corresponding nonzero coefficients ("nonnull variables"), while screening out variables with corresponding zero coefficients ("null variables"). This is commonly achieved by regularized regression using the Lasso penalty which returns a sparse estimator of β. In this case, the variables corresponding to nonzero estimated coefficients are returned as the nonnull variables. While asymptotic results show Lasso recovers the full set of nonnull variables, it also returns many null variables in practice, resulting in a high false discovery rate. Other developed algorithms, such as the knockoff filter (Barber and Candes 2015), conduct variable selection while ensuring the false discovery rate never exceeds a user-specified threshold. We propose a new procedure, "one-at-a-time knockoffs" (OATK), which relaxes the constraints of the knockoff filter and results in a more computationally efficient algorithm. Through simulations and real genetics data, we demonstrate OATK's higher power compared to competing methods and discuss OATK's theoretical properties to show that it also controls the false discovery rate.

2 - Optimal Inspection Policy under Imperfect Predictions: Structural Analysis and Insights

Guanlian Xiao, Vrije Universiteit Amsterdam, amsterdam, Netherlands, Alp Akcay, Lisa Maillart

We consider a critical component that deteriorates according to a three-state discrete-time Markov chain with a self-announcing failed state and two unobservable operational states: good and defective. The component is periodically monitored by a defect-prediction model that generates binary signals, but the signals are imperfect. The problem is to decide how to use the signals to make the inspect-or-not decision with the objective of minimizing the expected total discounted cost. We build a partially observable Markov decision process to address the problem.

3 - The General Symmetric Form of Linear Programs

Sion Ledbetter, SimpleRose, Saint Louis, MO, United States, Carl Ledbetter

There are several well known forms used to model linear programs, namely the canonical form, the standard form, and the general form. We present a novel formulation of linear programs, called the general symmetric form, which is a non-trivial extension of the general form. This extension is motivated by the fact that the dual of a general form linear program cannot be described in general form. In particular, the variable type which is dual to the bound variable type has never been characterized. We define and classify the dual of bound variables, called restricted variables, and demonstrate the necessity of objective coefficients for the logical variables defined by the constraints to generalize the dual decision variable bounds in general symmetric form. The symmetry between the primal and dual variable types and bounds in the general symmetric form provides the advantage of being able to construct a linear program from the primal and dual perspective simultaneously, giving modelers access to more tools and information with which to define their programs and resulting in a more accurate model. We establish and prove proper extensions of the canonical versions of Farkas' lemma, the duality equation, the existence-duality theorem, the weak duality theorem, the strong duality theorem, and the complementary slackness theorem for linear programs in general symmetric form.

4 - Peer Prediction for Verifier's Dilemma

Zishuo Zhao, University of Illinois Urbana-Champaign, Urbana, IL, United States

The security of blockchain systems is fundamentally based on the decentralized consensus in which the majority of parties behave honestly, hence the process of block verification is essential to keep the robustness of blockchain systems. However, the phenomenon that a secure blockchain system with few or no cheaters could not provide sufficient incentive for verifiers to honestly perform the costly verification, referred to as the Verifier's Dilemma, could severely undermine the fundamental security of blockchain systems. While existing works have attempted to insert deliberate errors to disincentivize lazy verification, the decentralized environment makes it impossible to directly judge the correctness of verification or detect malicious verifiers. In this work, we adopt the peer prediction method to design a Bayesian truthful mechanism for the verification game among multiple verifiers, incentivizing all verifiers to perform honest verification without access to the ground truth even in the presence of noisy verifications. With theoretically guaranteed truthfulness of our mechanism for the verification game, our work provides a framework of verification mechanisms that enhances the security and robustness of blockchain systems.

5 - Dynamic Time-Expanded Networks for Evacuation Planning

Angela Morrison, University of Colorado-Denver, Denver, CO, United States, Steffen Borgwardt, Nicholas Crawford, Drew Horton, Emily Speakman

We study the problem of evacuation planning for natural and human disasters, such as fire or venue evacuations. By creating pre-planned evacuation routes which can be updated based on real-time data, and by building on and pre-processing public data, we provide a holistic approach to fire evacuation planning and implementation. Evacuations can be formulated as max-flow problems on time-expanded networks. An initial flow is found based on a predicted fire, and then updated based on revised fire information at later time instances.We run test cases on three locations with historic deadly fires using data available through OpenStreetMaps on a NetworkX Python script. The results validate viable running times and quality of information for application in practice.

TD69

Regency - 607

SpORts I

Invited Session

spORts

Chair: Stephen Hill, Samford University, Birmingham, AL, United States

1 - Gaming the NET: Improving NCAA Tournament Outcomes with Scheduling

Stephen Hill, Samford University, Birmingham, AL, United States

Participation and placement in the the NCAA Men's Basketball Tournament is greatly affected by the NCAA's NET Rankings. In this presentation, the proprietary NET Ranking is reverse engineered and the effect of non-conference scheduling approaches on the NET is examined.

2 - Expected Points Above Average: A Novel NBA Player Metric Based on Bayesian Hierarchical Modeling

Ryan Elmore, University of Denver, Denver, CO, United States, Benjamin Williams, Erin Schliep, Bailey Fosdick

Team and player evaluation in professional sport is extremely important given the financial implications of success/failure. It is especially critical to identify and retain elite shooters in the National Basketball Association (NBA), one of the premier basketball leagues worldwide because the ultimate goal of the game is to score more points than one's opponent. To this end we propose two novel basketball metrics: “expected points” for team-based comparisons and “expected points above average (EPAA)” as a player-evaluation tool. Both metrics leverage posterior samples from Bayesian hierarchical modeling framework to cluster teams and players based on their shooting propensities and abilities. We illustrate the concepts for the top 100 shot takers over the last decade and offer our metric as an additional metric for evaluating players.

3 - College Football Survivor Pool

Eric Huggins, Katz School of Business, Fort Lewis College, Durango, CO, United States

NFL survivor pools are a game well-known to most fans of American football. Players must pick a different team to win each week for the entire season until they lose a game and are out; the last remaining “survivor” wins the pool. We discuss a related game, but using the college football Top 25 rather than the 32 NFL teams. This creates several twists: The Top 25 changes every week based on the previous week’s results – teams move in and out of the Top 25 and their rankings change from week to week. The pool also has multiple objectives which makes strategic decisions even more interesting.

I created this game about two decades ago and have been playing it ever since with friends, mostly for pride. Over time, it has evolved into a game with interesting and challenging decision-making that touches on many areas of analytics. With this in mind, I have started hosting the game in my Business Analytics class every Fall term – for fun and prizes, not for grades – although the students must discuss their decision-making processes and make calculations for class points.

In this talk, I will discuss the rules of the game, possible “optimal” strategies and if the timing works out, I will let the audience make my real picks for the following week! Further I will discuss how I use the game in class to reinforce understanding of various analytics topics.

4 - Go for It? Analyzing the Impact of Fourth Down Decisions in the NFL

Thomas Robbins, East Carolina University, Greenville, NC, United States

The decision to go for it on fourth down remains one of the most pivotal—and hotly debated—decisions an NFL coach can make. From Dan Campbell's perceived "reckless" choices to the Eagles' seemingly "unstoppable" tush push, these decisions have played a significant role in shaping the 2023-24 NFL season. This talk delves into the analytics behind fourth down decisions, presenting a standard framework for making the decision and an empirical analysis of game decisions and results. We will explore the alignment between coaches' decisions and analytical models, investigating the tangible effects of these decisions on game outcomes.

5 - Changing the Game: Analytics in Shaping Modern Sports On and Off the Field

Scott Nestler, Analytic Solutions, Granger, IN, United States

This presentation provides an overview of the transformative role of analytics in modern sports. It explores how data-driven approaches are revolutionizing performance on the field, enabling athletes and teams to optimize their strategies, enhance physical conditioning, and achieve superior results. Beyond the field, we examines the impact of analytics on the business side of sports, from talent scouting and player valuation to fan engagement and revenue generation. By integrating real-world examples and cutting-edge research, the presentation highlights the symbiotic relationship between analytics and sports, illustrating how data is not just reshaping strategies but also redefining the very essence of excellence in the sports industry.

TD70

Regency - 701

Electricity Storage

Invited Session

ENRE: Electricity

Chair: Dirk Lauinger, MIT, Cambridge

1 - Pricing Energy Storage for Social-Welfare Maximization

Ning Qi, Columbia University, New York, NY, United States, Ningkun Zheng, Bolun Xu

Existing market designs require energy storage to design bids strategically for profit maximization based on the future opportunity value. This benefits large and sophisticated storage participants who are experienced in price prediction and exercising market power. While, system operator and small storage participants unwilling to spend additional cost or intelligence to exercise market power will lose profit. Hence, this paper derives a novel theoretical framework to propose the opportunity pricing of energy storage for social welfare maximization. The opportunity price is formulated based on a two-stage chance-constrained framework, which guarantees the market equilibrium and social welfare maximization while incorporating uncertainties from renewables and load. We present tractable and deterministic reformulation for joint chance-constraints and different types of uncertainty realizations. Analytical results show that opportunity price decreases as the state of charge increases but rises with greater net load uncertainty. The key finding highlights the interconnected linear relationship between the opportunity price, electricity price, and reserve cost, demonstrating how energy storage can exercise market power in the energy and reserve market. We demonstrate the effecitveness of the proposed stochastic market design via the case study carried out on an 8-zone ISO-NE test system. Simulation results show that the proposed market design significantly reduces system costs and electricity payments, enhances decarbonization, and improves adaptability to uncertainty.

2 - Modeling Impacts of Multistage Operational Uncertainty on Long Duration Energy Storage

Thomas Lee, Massachusetts Institute of Technology, CAMBRIDGE, MA, United States, Andy Sun

Long duration energy storage may play a crucial role in highly decarbonized energy systems, in order to alleviate renewable energy shortages on multi-day and even seasonal timescales. However, out of system planning models to date which incorporate long duration storage, most assume perfect foresight during the operations of storage resources due to computational tractability. This conflicts with the limited skill of meteorological forecasts over climatology beyond the 1 to 2 week horizon, due to data/model limitations and inherent chaos in the weather system. The realism of this perfect foresight assumption is reexamined by considering a stochastic scenario tree based on historical weather data. Stochastic dual dynamic programming is used to study the behavior of long duration storage dispatch within an energy systems model under multistage operational uncertainty. Implications are studied for system planning, investment, and operations of energy storage portfolios with long
duration technologies.

3 - Arbitrage and Regulation with Electricity Storage

Dirk Lauinger, MIT, Cambridge, MA, United States, Andy Sun

Electricity storage is used for intertemporal price arbitrage and for balancing unforeseen fluctuations in electricity supply and demand, also known as regulation. We present an optimization model that computes bids for both arbitrage and regulation, and ensures that storage operators can honor their market commitments at all times for all fluctuation signals in an uncertainty set that is inspired by applicable legislation. We encode this requirement with an infinite number of functional constraints. We show that the robust constraints with functional uncertainties are equivalent to a finite number of deterministic constraints, which leads to an exact bilinear reformulation that can be solved with commercially available optimization software.

4 - Market Equilibria with Energy Storage as Flexibility Resources

Weiying Gu, The ohio state university, Columbus, OH, United States, Ramteen Sioshansi

Uncertain and variable real-time availability of renewable generation can increase the need for supply-side flexibility in power systems. Energy storage is a potential source of such flexibility. This paper examines the participation of multiple competing strategic profit-maximizing energy storage in a spot electricity market and its impact on consumers, producers, and market equilibria. To this end, we develop a two-stage stochastic bi-level model that has each energy-storage firm determine its market offers at the upper level to maximize its expected profit. The lower level represents market clearing under scenarios with different flexibility needs. We recast the bi-level model as a single-level optimization. A small illustrative example and a larger case study show that energy storage can increase market efficiency and reduce renewable-energy curtailment. We show that energy-storage firms neglecting uncertainty in optimizing their market offers can yield profit losses.

TD71

Regency - 702

Artificial Intelligence and Big Data Applications in Global Operations and Management

Invited Session

Data Mining

Chair: Pankaj Kumar, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Xiaojin Liu, Virginia Commonwealth University, Richmond, VA, United States

1 - Safe Adoption and Supervision of Generative AI Systems

Shubhranshu Singh, Johns Hopkins University, Baltimore, MD, United States

Generative AI systems often suffer from AI hallucinations in that they may provide confident responses unjustified by their training data. Given the increasing reliance on generative AI systems, it is hard to overstate the damage that AI hallucinations may engender. In this paper, we explore the question of whether AI developers or human decision makers should be liable for the harm the AI systems may cause. We develop a novel theoretical framework to study and compare implications of liability for developers of AI systems versus liability for human decision makers. We find that in some cases AI developer liability can interestingly result in over-supervision and under-adoption by the human decision makers. Although AI operator liability leads to efficient accuracy investment when AI accuracy is publicly observed, developer liability achieves efficient accuracy investment when it is privately observed by the developer.

2 - Navigating the Communication Divide: Using Large Language Models for Enhanced Patient-Provider Interaction in Opioid Disorder Treatment

Murtaza Nasir, Barton School of Business, Wichita State University, Wichita, KS, United States, Anton Ivanov, Jasmina Tacheva

<p break-words="">The opioid epidemic presents significant challenges to public health, and communication issues often arise when patients with opioid use disorder (OUD) interact with healthcare providers. The use of stigmatized language by clinicians can hinder effective dialogue and treatment plans. Adaptive AI-based systems utilizing large language models (LLMs) could potentially provide real-time communication support to practitioners, but the quality of the generated feedback is crucial, especially when dealing with sensitive populations like OUD patients.

<p break-words="">This study investigates whether LLMs can improve patient-provider communication by reducing stigmatized language and enhancing health outcomes. Adapting the Communication Accommodation Theory (CAT), the researchers collected a dataset of conversations and health outcomes for 192 OUD patients from an online health community on Reddit. They benchmarked several LLMs and compared their performance in reducing stigmatized language, using survival analysis to examine the effects of machine- vs. human-generated content on patient survival in terms of sobriety and relapse.

<p break-words="">The results suggest that machine-generated content lacking stigmatized language is associated with a 17% increase in the probability of a patient reaching a subsequent clean stage and a 9% reduction in the probability of relapse. The study highlights the potential of LLMs to enhance communication in healthcare settings and support patients across various domains.

3 - Text-data Mining of Ridesharing Firms: An Integrated NLP-MCDM Approach

Arnab Bisi, Johns Hopkins Carey Business School, Baltimore, MD, United States, Souradeep Koley, Mukesh Barua

The past decade witnessed an upsurge in the operation of the Transportation Network Companies (TNCs) like Uber and Lyft on a global level. Leveraging online platforms and mobile technology, these organizations have a substantial influence on travelers, transportation systems, and international communities. While existing literature centered on passenger adoption and drivers' motivation, our study addresses identifying the challenges faced by these ridesharing firms. We use large-scale publicly available data from social media to gain an initial understanding of consumer issues related to organizations. We use a machine learning approach to identify the latent themes, followed by emotion analysis to understand the sentiments. The findings highlight issues related to *'legal activity and regulatory compliance'* and *'service availability'* among consumers from the ridesharing firms. Although most riders have positive sentiments (34.13%), there is a relative increase in anger and fear among the consumers. Following these, empirically, using decision-making, we identify and quantitatively address five major inhibitors faced by these organizations. Contrary to conventional perception, our study highlights the *'government'* as a major inhibitor due to stringent government regulations and the discrepancy between central and state. Additionally, the negative word-of-mouth and network effects have created significant issues for ridesharing firms over the years. The findings from the integrated methods provide multiple insights for the policymakers and managers in TNCs to increase the overall efficiency of the firms.

TD72

Regency - 703

Advances in Data-Driven Decision-Making in Operations Management

Invited Session

Data Mining

Chair: Yonggab Kim, Purdue University, West Lafayette, IN, United States

1 - Developing Protocols for Autonomous Mobile Robots in Material Handling Using an Inductive Learning Approach

Byeongmok Kim, Purdue University, West Lafayette, IN, United States

In a manufacturing environment, autonomous mobile robots (AMRs) operate autonomously and with greater operational flexibility than automated guided vehicles (AGVs), which are controlled by a central system. On the other hand, the inability to find an operationally optimal solution from a global perspective partially offsets the operational efficiency of AMRs. In our study, we propose a new learning framework to achieve both operational efficiency and flexibility of operations of AMRs by learning the operating policies of AMRs based on the optimal solution of a centralized system.

2 - Enhancing Genetic Algorithm with Explainable AI for Last-Mile Routing

Yonggab Kim, Purdue University, West Lafayette, IN, United States, Reem Khir, Seokcheon Lee

Traditional evolutionary optimization algorithms often presuppose straightforward constraints and objective functions. However, in many real-world optimization problems, a clear-cut objective function may be absent or hard to evaluate. Given these challenges, surrogate models have gained attention as proxies for evaluating objectives or constraints. Surrogate models, when integrated into Evolutionary Algorithms (EAs), result in Surrogate-Assisted EAs (SAEAs). In this research, we leverage machine learning and Explainable Artificial Intelligence (XAI) to enhance the performance of Genetic Algorithms (GAs) for combinatorial optimization. Our framework utilizes a surrogate model to capture the nuanced interplay within environmental data. We interpret these models using SHapley Additive exPlanations (SHAP) values, identifying key features for prioritization. This understanding enables us to enhance the design of the genetic algorithm, covering the development of its structure and operators, as well as the calibration of fitness functions. Such tailored adjustments enhance the GA's efficiency and effectiveness. We demonstrate our framework on a sequencing problem that arise in last-mile routing using data from the Amazon 2021 Last Mile Challenge. Our methodology extracts drivers' preferred routing characteristics and applies these insights to generate new routes. Our method achieved a 0.0411 score, ranking fourth and demonstrating improvements in the convergence and effectiveness of GAs.

3 - Generative AI-Driven Socio-Economic Agent-based Simulation

Seung ho woo, Purdue University, West Lafayette, IN, United States, Seonho Woo, Young-Jun Son

Contemporary society faces a range of disruptions of natural or human-made crises, such as pandemics, hurricanes, and cyber-attacks that destabilize social and economic system dynamics. To enhance the resilience of our society under various disruptions, predicting socio-economic outcomes utilizing a high-fidelity agent-based socio-economic simulation is essential. However, constructing reliable agent-based simulations is costly, time-consuming, and complex. This novel research pioneers high-fidelity socio-economic prediction by integrating generative AI with agent-based simulation. This study employs Generative Artificial Intelligence (AI) to construct realistic environments and develop human-like agents for high accuracy economic simulations. The aim is to predict societal outcomes accurately under various disruptions, providing a practical tool for policymakers and economists. Generative models are applied to create a small city with diverse infrastructures that resemble real-world settings. Additionally, reinforcement learning-induced generative models are used to create human-like agents with realistic purchasing behaviors, reflecting the complex social phenomena of the real world. Our generative agent-based simulation framework aims to generate multi-agent societal system simulations with continuously learning agents in a generated environment. This study simulates small town-level multi-agent diverse scenarios to observe market dynamics, including market trends with supply and stock level assessment under disruption, by applying the potential of a generative AI model.

4 - Integrated Dynamic Rebalancing and Charging for Electric Scooter Sharing System

Zhuoli Yin, Purdue University, West Lafayette, IN, United States, Reem Khir, Hua Cai

Electric-Scooter Sharing System (ESS) is a popular dock-less micro-mobility option for first and last mile trips in cities. ESSs can reduce traffic congestion and lower greenhouse gas emissions by replacing private cars and enhancing public transit usage. Because of spatiotemporally imbalanced customer demands and limited battery ranges, ESSs face operational challenges in sustaining their functionality throughout the day. Customers unable to find available, sufficiently charged e-scooters are likely to leave the system, resulting in unsatisfactory service experiences. To address this issue, providers must regularly rebalance e-scooters across zones and return them to depots for charging. However, prior studies on rebalancing and charging of ESS are limited. Research has primarily focused on isolated strategies for either rebalancing or charging, or on overnight operations that do not account for dynamic daytime customer demands. In this study, we proposed an integrated dynamic rebalancing and charging model based on mixed integer programming. This model considers both the real-time battery levels of e-scooters and their distribution across service zones and depots. To efficiently address real-world, large-scale instances, we applied Lagrangian Dual Decomposition and Local Search methods to approximate solutions, breaking the problem into three subproblems: rebalancing vehicle routing, e-scooter allocation, and charging restoration. Our numerical experiments demonstrate that our model generates high-quality solutions for large-scale instances. Our model offers an approach for dynamically maintaining the service of shared mobility systems with electrified fleets.

TD73

Regency - 704

Recent Advances in Multi-Agent and Preference-Based Reinforcement Learning

Invited Session

Data Mining

Chair: Junzi Zhang, Citadel Securities, New York, NY, 60654, United States

1 - MF-OML: Online Mean-Field Reinforcement Learning with Occupation Measures for Large Population Games

Anran Hu, Oxford University, Oxford, United Kingdom

Reinforcement learning for multi-agent games has attracted lots of attention recently. However, given the challenge of solving Nash equilibria for large population games, existing works with guaranteed polynomial complexities either focus on variants of zero-sum and potential games, or aim at solving (coarse) correlated equilibria, or require access to simulators, or rely on certain assumptions that are hard to verify. This work proposes \texttt{MF-OML} (Mean-Field Occupation-Measure Learning), an online mean-field reinforcement learning algorithm for computing approximate Nash equilibria of large population sequential symmetric games. \texttt{MF-OML} is the first fully polynomial multi-agent reinforcement learning algorithm for provably solving Nash equilibria (up to mean-field approximation gaps that vanish as the number of players N goes to infinity) beyond variants of zero-sum and potential games. When evaluated by the cumulative deviation from Nash equilibria, the algorithm is shown to achieve a high probability regret bound of $\tilde{O}(M^{3/4}+N^{-1/2}M)$ for games with the strong Lasry-Lions monotonicity condition, and a regret bound of $\tilde{O}(M^{11/12}+N^{-1/6}M)$ for games with only the Lasry-Lions monotonicity condition, where M is the total number of episodes and N is the number of agents of the game. As a by-product, we also obtain the first tractable globally convergent computational algorithm for computing approximate Nash equilibria of monotone mean-field games.

2 - Beyond Equilibrium Learning

Chi Jin, Princeton University, Princeton, NJ, United States

While classical game theory primarily focuses on finding equilibria, modern machine learning applications introduce a series of new challenges where standard equilibrium notions are no longer sufficient, and the development of new efficient algorithmic solutions is urgently needed. In this talk, we will demonstrate two such scenarios: (1) a natural goal in multiagent learning is to learn rationalizable behavior, which avoids iteratively dominated actions. Unfortunately, such rationalizability is not guaranteed by standard equilibria, especially when approximation errors are present. Our work presents the first line of efficient algorithms for learning rationalizable equilibria with sample complexities that are polynomial in all problem parameters, including the number of players; (2) In multiplayer symmetric constant-sum games like Mahjong or Poker, a natural baseline is to achieve an equal share of the total reward. We demonstrate that the self-play meta-algorithms used by existing state-of-the-art systems can fail to achieve this simple baseline in general symmetric games. We will then discuss the new principled solution concept required to achieve this goal.

3 - Actions Speak What You Want: Provably Sample-Efficient Reinforcement Learning of the Quantal Stackelberg Equilibrium from Strategic Feedbacks

Siyu Chen, Yale University, New Haven, CT, United States, Mengdi Wang, Zhuoran Yang

We study reinforcement learning (RL) for learning a Quantal Stackelberg Equilibrium (QSE) in an episodic Markov game with a leader-follower structure. In specific, at the outset of the game, the leader announces her policy to the follower and commits to it. The follower observes the leader's policy and, in turn, adopts a quantal response policy by solving an entropy-regularized policy optimization problem induced by leader's policy. The goal of the leader is to find her optimal policy, which yields the optimal expected total return, by interacting with the follower and learning from data. A key challenge of this problem is that the leader cannot observe the follower's reward, and needs to infer the follower's quantal response model from his actions against leader's policies. We propose sample-efficient algorithms for both the online and offline settings, in the context of function approximation. Our algorithms are based on (i) learning the quantal response model via maximum likelihood estimation and (ii) model-free or model-based RL for solving the leader's decision making problem, and we show that they achieve sublinear regret upper bounds. Moreover, we quantify the uncertainty of these estimators and leverage the uncertainty to implement optimistic and pessimistic algorithms for online and offline settings. Besides, when specialized to the linear and myopic setting, our algorithms are also computationally efficient. Our theoretical analysis features a novel performance-difference lemma which incorporates the error of quantal response model, which might be of independent interest.

TD74

Regency - 705

Net-Zero Emissions Energy Systems: Optimization Methods

Invited Session

ENRE: Energy-Climate

Chair: Edgar Virgüez, Carnegie Science, Stanford, CA, United States

Co-Chair: Michael Lau, Princeton University ZERO Lab, Princeton, NJ, United States

Co-Chair: Aleksander Grochowicz, University of Oslo, Oslo, Norway

1 - Funplex: A Modified Simplex Algorithm to Efficiently Explore Near-Optimal Spaces

Linda Brodnicke, ETH Zurich, Zurich, Switzerland, Christoph Funke, Francesco Lombardi, Giovanni Sansavini

Modeling to generate alternatives (MGA) is an increasingly popular method in energy system optimization. MGA explores the near-optimal space, namely, system alternatives whose costs are within a certain fraction of the globally optimal cost. Real-world stakeholders may prefer these alternatives due to intangible factors. Nonetheless, widespread MGA adoption is hampered by its additional computational burden. Current MGA methods identify boundary points of the near-optimal space through repeated, independent optimization problems. Hundreds of model runs are usually required, and such individual runs are often inefficient because they repeat calculations or retrace previous trajectories. In this study, we transcend such limitations by introducing a novel algorithm called Funplex, which uses methods from multi-objective Simplex to optimize many MGA objectives with minimal computational redundancy. For a simple linear-programming energy hub case study, we show that Funplex is ten times faster than existing methods and yields higher-quality near-optimal spaces. Furthermore, sensitivity analyses suggest that Funplex scales well with the number of investment variables, making it promising for capacity planning models. The current proof-of-concept implementation based on a full multi-objective tableau may face memory and stability limitations for large models. Nonetheless, future developments based on more advanced versions of Simplex may overcome such barriers, thereby making MGA more accessible and standard among modeling teams.

2 - A Computationally Efficient Modelling to Generate Alternatives Algorithm using Benders Decomposition

Michael Lau, Princeton University, Princeton, NJ, United States, Filippo Pecci, Anna Jacobson, Jesse Jenkins

Contemporary macro-energy systems modelling is characterized by the need to represent energy systems strategic and operational decisions with high temporal and spatial resolution, which provides more accurate results than more abstracted models. This drive towards greater fidelity, however, conflicts with a push towards greater model representation of inherent complexity in decision-making, including methods like Modelling to Generate Alternatives. Modelling to Generate Alternatives aims to map the feasible space of a model within a cost slack by varying investment parameters without changing the operational constraints, a process which frequently requires hundreds of solutions. For large, highly representative energy system models, solving for large numbers of highly detailed solutions is impossible with traditional methods, leading researchers to reduce complexity with through zonal or temporal aggregation. This research presents a new solution method for Modelling to Generate Alternatives-type problems. Using Benders Decomposition, we break down the problem structure into a strategic master problem and operational subproblems and pass information between master problems to accelerate convergence with each new objective. We find that our new solution method is several times faster and requires less memory than existing parallelized monolithic Modelling to Generate Alternatives solution methods, enabling rapid computation of a greater number of solutions to highly resolved models.

3 - Charging Fleets of Electric Vehicles at Airport Rental Car Centers

Devon Sigler, National Renewable Energy Laboratory, Crested Butte, CO, United States

At major airports hundreds of vehicles are returned to the rental car center every hour. These vehicles must be cleaned and refueled to be rented out again. As rental car fleets transition to electric vehicles (EV), naïve first come first serve charging strategies have the potential to create peak power loads in the tens of megawatts range, which existing airport infrastructure may not be equipped to handle. However, through strategic scheduling of charging start times and charging power levels these peak loads can be mitigated to not over stress existing infrastructure. In this talk we present work on a column generation algorithm for solving this EV scheduling problem at realistic scales fast enough to be practically useful.

4 - Diverse developments of European green hydrogen production

Koen van Greevenbroek, UiT The Arctic University of Norway, Tromsø, Norway, Johannes Schmidt, Marianne Zeyringer, Alexander Horsch

Green hydrogen is seen by the EU as a key technology in the energy transition towards carbon net neutrality; the Commission has committed to domestic production and imports of 10Mt (~330TWh) each by 2030. The prospects for green hydrogen production in the following decades, however, is the subject of much debate. While domestic green hydrogen could improve energy security and enable the switch away from fossil fuels, it is dependent on ample renewable energy sources and may face competition from oil and natural gas combined with carbon capture and storage, as well as imports from outside of Europe.

We introduce a novel way of exploring different development pathways for a single resource such as green hydrogen. Our work builds on ideas from the Modelling to Generate Alternatives (MGA) framework, which can be used to find near-optimal system designs. By applying MGA-type steps at each point in a sequence of optimisations modelling the evolution of the European energy system from 2025 to 2050, we reveal a wide variety of feasible scenarios for the growth of green hydrogen production. While a moderate domestic production level is often cost-optimal, we find that anything from large production volumes to no production at all can be accommodated for under most assumptions with only a minor total system cost increase. Near-optimal methods thus indicate that there is ample space for social and political considerations in cost-effective green hydrogen policy.

5 - Exploring Robust pathways for Incentivizing DAC deployment across power systems under uncertainty

Franklyn Kanyako, University of Michigan, Ann Arbor, MI, United States, Michael Craig

Direct Air Capture (DAC) technology has the potential to play a crucial role in achieving negative emissions, but its widespread deployment depends on effective public policy support. Designing policies that accelerate DAC adoption while hedging against premature investment and long-term delays is challenging due to inherent uncertainties in technology costs, mitigation policies, timelines, and the availability of alternative low-carbon energy sources. The combination of multiple policy instruments and technical pathways further complicates decision-making for policymakers, as the effects of specific policies on future DAC deployment are highly uncertain. Limited research exists on policy intervention and investment needs to accelerate DAC deployment robust to near-term and long-term uncertainties. This examines the role of R&D subsidies and climate policy in accelerating the near-term and long-term deployment of DAC across the United States electric power system. By developing a comprehensive framework that incorporates multiple uncertainties and policy instruments, this research aims to provide policymakers with insights into designing effective and robust policies for accelerating DAC deployment while managing the risks associated with premature investment and long-term delays.

TD75

Regency - 706

Analysis Toward Net-Zero Transportation

Invited Session

ENRE: Environment and Sustainability

Chair: Wenquan Dong, Texas State University, San Marcos, TX, 78666, United States

1 - On the mathematical modeling of the multi-depot green vehicle routing problem with multiple one-to-many pickup and delivery operations and time windows

Alejandro Fernandez Gil, Texas State University, San Marcos, TX, United States, Eduardo Lalla-Ruiz, Martijn Mes, Wenquan Dong

This work focuses on freight distribution and introduces the multi-depot green vehicle routing problem with pickup and delivery operations and time windows (MDGVRP-PD-TW), aiming to minimize emissions through reduced fuel consumption. A mathematical formulation is proposed that respects the time windows established for each customer and considering multiple one-to-many pickups and deliveries, where each pickup customer is associated with a set of delivery customers. By including the sumproduct between flow and distance traveled over the arcs, routing decisions incorporate the environmental impact of CO2 emissions, allowing for a balance between emissions and weight transported along the routes. To solve this problem, we propose a matheuristic approach based on the POPMUSIC framework (Partial Optimization Metaheuristics under Special Intensification Conditions). The proposed solution approaches are tested on a new benchmark suite designed to evaluate this variant.

2 - Developing Sustainable Transportation with Enhanced System Efficiency and ITS Technologies

Fengxiang Qiao, Texas Southern University, Houston, TX, United States, Jianbang Du, Qing Li

The United States Bipartisan Infrastructure Law supports the development of sustainable transportation with low- to zero-emission, energy-efficient, and affordable transportation modes. The benefits include (1) savings on fuel and vehicle costs, (2) less carbon and toxic emissions and less air pollution, (3) increased economic productivity (e.g., more job opportunities in new energy and vehicle productions), (4) more reliable and affordable transportation with improved travel equality and environmental justice, and (5) enhanced security and independence. Sustainable transportation involves efforts such as developing new energy and fuel, managing travel demand and mode shift, and integrating equity, justice, and community engagement. In this presentation, the presenters wish to share several examples of sustainable transportation strategies with enhanced transportation efficiency and the technologies related to Intelligent Transportation Systems (ITS). Typical exemplary strategies include (1) improving congestion-pricing strategies for reduced emissions (e.g., ozone) along freeway corridors with High-Occupancy Vehicle (HOV) / High-Occupancy Toll (HOT) lanes, (2) improving freeway ramp metering strategies for reduced environmental impacts, (3) developing reinforced learning-based eco-transportation strategies at intersections and transportation networks with minimized ecological impacts, (4) improved car-following strategies along an arterial street with multiple coordinated signalized intersections, (5) characterizing roadway pavement roughness impacts on vehicle emissions with minimized air pollution, and (6) improving roadway design (e.g., the length and type of freeway weaving area) with less impacts on vehicle emissions and drivers' health. It is recommended that advanced technologies such as artificial intelligence, machine learning, and deep learning be employed more systematically for more environmentally friendly transportation systems.

3 - Locating and Allocating Battery Electric Bus Chargers with Stochastic Charging Demand

Sadjad Bazarnovi, University of Illinois at Chicago, Chicago, IL, United States, Taner Cokyasar, Omer Verbas, Abolfazl (Kouros) Mohammadian

Bus electrification accelerates the transition towards sustainable urban transportation. Battery Electric Buses (BEBs) often require periodic recharging intervals between service trips. The strategic placement and allocation of charging stations are crucial in facilitating operations, optimizing costs, and minimizing downtime. In this study, we develop a model, the Stochastic ELECTRic vehicle charger Allocation (SELECTRA) framework, to address the optimal distribution of various plug types at candidate locations under the stochastic charging demand of BEBs. Our strategic planning framework aims to minimize long-term system costs while balancing considerations among station, charger, and value of time expenses. We develop a mixed-integer non-linear program (MINLP) to solve this model. Leveraging existing stochastic location literature, we use an exact solution method to address this computationally challenging problem. The program optimizes the generalized cost associated with building charging stations, allocating charger types, travel to charging stations, and average queueing and charging times. Queueing dynamics are modeled using an M/M/s queue, with the number of servers as a decision variable. The linearization process, which excludes queueing variables, is succeeded by introducing associated constraints through cutting planes. This step guarantees global optimality due to the convexity properties of the problem. To enhance scalability, we implement a simulated annealing metaheuristic and devise heuristic clustering strategies, facilitating the effective resolution of real-world, large-scale instances. We conduct comparative analyses across garage-only, other-only, and mixed-location scenarios, along with sensitivity analyses on key parameters.

TD76

Regency - 707

Energy System Modeling and Optimization V: ML-based Multi-Energy System Optimization

Invited Session

ENRE: Other Energy

Chair: Qifeng Li, University of Central Florida, Orlando, FL, United States

1 - An ML-accelerated solution method for real-time optimization of energy-water-hydrogen nexus

Qifeng Li, University of Central Florida, Orlando, FL, United States

The micro Energy-Water-Hydrogen (m-EWH) nexus, which utilizes renewable energy sources (RES) to produce hydrogen by electrolysis, which is then combined with carbon captured from fossil fuel power plants, is investigated for reducing carbon emissions from the power sector. To address RES uncertainty, this paper proposes a real-time decision-making scheme for the m-EWH nexus that requires quickly solving large-scale mixed-integer convex programming (MICP) problems. For this purpose, a machine learning-accelerated solution method for real-time optimization (MARO) consisting of three modules is developed. First, an active constraint and integer variable prediction approach is introduced to rapidly solve MICP problems using historical optimization data. Second, a feasible rank-optimal strategy selection module is proposed to ensure the feasibility of the obtained solutions. Last but not least, a feature space extension module is developed to improve the accuracy of obtained solutions by creating new features and refining existing ones.

2 - Enhancing Resilience of Grid-Interactive Efficient Buildings Against Heat Waves Using Reinforcement Learning

Chenxi Hu, The University of Hong Kong, Hong kong, China, People's Republic of, Yujia Li, Yixuan Chen, Yunhe Hou

Heat waves, characterized by abnormally high temperatures, can significantly increase electricity demand within Grid-Interactive Efficient Building (GEB) clusters, challenging grid power capacity and potentially leading to failures in cooling and power supply. This study explores the resilience of GEBs during extreme heat waves using reinforcement learning. To achieve resilient operation under heat wave conditions, this study aims to find the optimal control policy for multiple flexible resources within GEB clusters. We employ the Proximal Policy Optimization (PPO) algorithm to develop an adaptive control strategy that dynamically responds to changing conditions and optimizes the use of available resources. The performance of the PPO-based method is compared against rule-based controls to assess improvements in operational resilience. Preliminary results demonstrate that the PPO-based RL approach enhances the resilience of GEBs by effectively managing electricity usage and maintaining thermal comfort, even when grid power is restricted. This study underscores the potential of RL in developing robust, adaptive control policies for GEB clusters, offering a promising method for mitigating the impacts of extreme heat waves on urban energy systems.

3 - ML-based optimization for accelerating complex grid reliability assessment under extreme weather events

Qiuhua Huang, Colorado School of Mines, Golden, CO, United States, Xiaolin Chen

In the face of increasing extreme weather events, the resilience and reliability of electrical grids are paramount. This talk explores the application of machine learning (ML) techniques to optimize and accelerate the optimization module within the grid reliability assessment process under such conditions. Traditional methods for evaluating grid reliability can be computationally intensive and slow, especially when simulating complex scenarios influenced by severe weather due to large number of scenarios involved. By leveraging advanced ML algorithms, we can significantly enhance the speed of these assessments with good accuracy. This presentation will cover the development and integration of ML models for short-term power system reliability assessment, the challenges encountered, and the solutions implemented.

4 - Physics-Informed Machine Learning to Enhance Electric Distribution System Situational Awareness

Ying Zhang, Okalahoma State university, Stillwater, OK, United States

Modern distribution systems are confronted with degraded and even stranded operational performances subject to growing complexity, uncertainty, and volatility from large-scale integration of renewable and distributed energy resources (DERs). This talk brings new perspectives to learning-to-optimize (L2O) distribution system operation by focusing on fast yet accurate nonlinear power flow calculation with DERs–combining power engineering knowledge with data-aided machine learning techniques, i.e., physics-informed machine learning – to replace a blind application of "black-box" data-intensive ones, and the latter has limited practicality in real-world grids. The proposed physics-informed L2O method achieves significant performance synergy in terms of accuracy, calculation time, topology adaptivity, and robustness against data errors, outliers, or cyberattacked measurements. As power flow calculation is of fundamental importance for grid monitoring, control, and management, the proposed framework has the potential as a building block to accelerate and advance various optimization tasks to improve the situational awareness towards the next-generation distribution grids.

TD77

Regency - 708

Advances in Decomposition Algorithms

Invited Session

Computing Society

Chair: Geunyeong Byeon, Arizona State University, Tempe, AZ, United States

1 - Benders Lift-and-Project Cuts

Kaiwen Fang, Arizona State University, Phoenix, AZ, United States, Geunyeong Byeon

In this paper, we explore the utilization of disjunctive system to enhance the strength of cuts within a Benders decomposition algorithm. At a broad level, the proposed algorithm leverages integrality information and Benders cuts to generate cuts tightly aligned with the convex hull of the Benders reformulation. Importantly, this is achieved via iteration of standard Benders subproblem. Specifically, our study highlights that normalization plays a crucial role in eliminating the generation of overly weak cuts. Meanwhile, we found that utilizing the cut-and-project technique can make the generated disjunctive benders cuts more effective. We conduct a comparative analysis to assess the impact of Benders disjunctive cuts under various normalization settings in facility location problems.

2 - Characterizing and Learning the Lagrangian Multipliers for the Bin Packing Problem

Yoonju Sim, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, Republic of, Hyeonah Kim, Changhyun Kwon, Sungbin Lim

We consider the Lagrangian relaxation method for finding a dual bound for the Bin Packing Problem. A tight dual bound plays a critical role in closing the gap for any exact algorithm. Furthermore, in the context of vehicle routing problems, all capacity inequalities require fast computation of a good dual bound. In this paper, we study the characteristics of optimal Lagrangian dual solutions such as the relationship between the primal and dual solutions, dual polyhedron, and so on. By using them, we aim to speed up the subgradient optimization method used in the Lagrangian relaxation method.

3 - The Chance-Constrained Stochastic Diversion Path Problem with Sample Average Approximation

Woojin Kim, University of Wisconsin-Madison, Madison, WI, United States, Jim Luedtke, Cole Smith

The diversion path problem is a network interdiction problem in which the leader modifies arc lengths, and the follower finds the shortest path from the source node to the sink node using the modified arc lengths. The leader's goal is to minimize the cost of having the follower choose a special path called the diversion path. When arc lengths are deterministic, this problem is solvable by linear programming. We study the situation where arc lengths are random, and the leader wishes to minimize the cost to ensure the follower follows the diversion path with the probability of at least a desired target. This chance-constrained stochastic diversion path problem has been studied by Baycik, Nguyen, and Smith in the special case when arc lengths are independent and normally distributed. We explore the use of sample average approximation (SAA) to approximately solve this problem with a general distribution. We first derive a compact mixed integer programming formulation of the SAA problem but find that it is challenging to solve due to poor linear programming relaxation quality. We thus explore a method based on a branch-and-cut decomposition algorithm. Computational results will be presented.

4 - A Gpu-Based Distributed Algorithm for Linearized Optimal Power Flow in Distribution Systems

Minseok Ryu, Arizona State University, Tempe, AZ, United States, Geunyeong Byeon, Kibaek Kim

We propose a GPU-based distributed algorithm, aimed at controlling optimal power flow in distribution systems. Typically, conventional algorithms employed in such scenarios rely on parallel execution with multiple CPUs. However, this often leads to significant computation time primarily due to the need for optimization solvers to solve subproblems. To address this computational challenge, we propose a distributed algorithm that eliminates solver dependencies and harnesses GPU acceleration. We demonstrate the computational performance of our approach through numerical experiments on four IEEE test instances ranging from 13 to 8500 buses. Our results affirm the scalability and superior speed of our GPU-based approach compared to the CPU-based counterpart.

TD78

Regency - 709

Federated Learning and Optimization: II

Invited Session

Computing Society

Chair: Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

1 - Asynchronous Federated Stochastic Optimization for Heterogeneous Local Objectives under Arbitrary Delays

Charikleia Iakovidou, Argonne National Laboratory, Lemont, IL, United States, Kibaek Kim

Federated learning (FL) was recently proposed to securely train models with data held over multiple locations ("clients") under the coordination of a central server. Two major challenges hindering the performance of FL algorithms are long training times caused by straggling clients and a decrease in training accuracy induced by non-iid local distributions, modeled as heterogeneous local functions. To address these limitations, we propose a new stochastic (sub)gradient algorithm that is robust to client drift and utilizes asynchronous communication to speed up convergence in the presence of stragglers. Moreover, the proposed algorithm is, to the best of our knowledge, the first method that is both guaranteed to converge under arbitrarily long delays, and converges to an error neighborhood whose size depends only on the variance of the stochastic (sub)gradients used and thus is independent of both the heterogeneity between the local datasets, and the length of client delays, without the use of delay-adaptive stepsizes. Our numerical results confirm our theoretical analysis and suggest that the proposed method outperforms the state-of-the-art when local datasets are highly non-iid.

2 - Federated low-rank image reconstruction with multimodal CT

Julie Anh Nguyen, Northwestern University, Evanston, IL, United States, Zichao Di, Diego Klabjan, Minseok Ryu, Kibaek Kim

Tomographic image reconstruction is an important scientific problem that has significant applications in many fields. Tomographic projections in many cases are of large volumes and/or three-dimensional. Previous works have shown that joint reconstruction using data from different modalities can alleviate the ill-posedness of the single-modality problem, which is typically due to insufficient measurements and noise. In addition, exploiting the underlying tomographic structure through low-rank approximation of the object of interest has shown promising results, especially with limited-angle projections. Within the federated framework in which clients with different modalities collaborate and perform joint image reconstruction, we develop an algorithm that leverages low-rank approximation to not only capture latent multilinear structure but also improve communication efficiency.

3 - An Online Optimization Perspective on First-Order and Zero-Order Decentralized Nonsmooth Nonconvex Stochastic Optimization

Muhammet Emre Sahinoglu, Northeastern University, Boston, MA, United States, Shahin Shahrampour

We investigate the finite-time analysis of finding (δ,ϵ) stationary points for nonsmooth nonconvex objectives in decentralized stochastic optimization. A set of agents aim at minimizing a global function using only their local information by interacting over a network. We present a novel algorithm, called Multi Epoch Decentralized Online Learning (ME-DOL), for which we establish the sample complexity in various settings. First, using a recently proposed online-to-nonconvex technique, we show that our algorithm recovers the optimal convergence rate of smooth nonconvex objectives. We further extend our analysis to the nonsmooth setting, building on properties of randomized smoothing and Goldstein-subdifferential sets. We establish the rate of $O(\delta^{-1}\epsilon^{-3})$, which to the best of our knowledge is the first finite-time guarantee for general decentralized nonsmooth nonconvex objectives in the first-order oracle setting, matching its optimal centralized counterpart. We further prove the same rate for the zero-order oracle setting without using variance reduction.

4 - Improved guarantees for optimal Nash equilibrium seeking and bilevel variational inequalities

Sepideh Samadigharahbalagh, Rutgers university, Piscataway, NJ, United States, Farzad Yousefian

We consider a class of hierarchical variational inequality (VI) problems that subsumes VI-constrained optimization and several other important problem classes including the optimal solution selection problem and the optimal Nash equilibrium (NE) seeking problem. Our main contributions are threefold. (i) We consider bilevel VIs with monotone and Lipschitz continuous mappings and devise a single-timescale iteratively regularized extragradient method, named IR-EG(m,m). We improve the existing iteration complexity results for addressing both bilevel VI and VI-constrained convex optimization problems. (ii) Under the strong monotonicity of the outer level mapping, we develop a method named IR-EG(s,m) and derive faster guarantees than those in (i). We also study the iteration complexity of this method under a constant regularization parameter. These results appear to be new for both bilevel VIs and VI-constrained optimization. (iii) To our knowledge, complexity guarantees for computing the optimal NE in nonconvex settings do not exist. Motivated by this lacuna, we consider VI-constrained nonconvex optimization problems and devise an inexactly-projected gradient method, named IPR-EG, where the projection onto the unknown set of equilibria is performed using IR-EG(s,m) with a prescribed termination criterion and an adaptive regularization parameter. We obtain new complexity guarantees in terms of a residual map and an infeasibility metric for computing a stationary point. We validate the theoretical findings using preliminary numerical experiments for computing the best and the worst Nash equilibria.

Tuesday, October 22, 4:00 PM - 5:15 PM

TE01

Summit - 320

Explainable AI for prognostics

Invited Session

Quality, Statistics and Reliability

Chair: Ye Kwon Huh, University of Wisconsin-Madison, Madison, WI, 53705, United States

Co-Chair: Kaibo Liu, UW-Madison, Madison, WI, United States

1 - Strata design for variance reduction in stochastic simulation

Jaeshin Park, University of Michigan, Ann Arbor, Ann arbor, MI, United States, Young Myoung Ko, Sara Shashaani, Eunshin Byon

Stratified sampling is one of the powerful variance reduction methods for analyzing system performance with stochastic simulation. However, strata design faces the curse of dimensionality and data scarcity as the input dimension increases. We analytically derive the optimal stratification structure that minimizes the estimation variance for univariate problems. Further, reconciling the optimal stratification into decision trees, we devise a robust algorithm for multi-dimensional problems.

2 - An uncertainty-informed neural network-based prognostic model for multi-type data

Ye Kwon Huh, University of Wisconsin-Madison, Madison, WI, United States

Modern prognostic systems generate multi-type data, typically in the form of event logs with discrete event data and signal records with continuous signal data. Although each data type provides unique insights into the degradation status of the system, existing prognostic methods are limited as they focus solely on either data type to predict the remaining useful life (RUL). To fill this gap, we propose a novel neural network-based model that can harness prognostic insights from both discrete event data and continuous signal data. Specifically, the proposed model contains an event predictor for the event time and types, a signal predictor for the degradation signals, and an RUL predictor that integrates the signal and event predictions to ultimately predict the RUL. To prevent unwanted bias and under/overfitting issues during model training, the proposed model adopts an uncertainty-informed joint training process that leverages uncertainty information to automatically learn the relative weights of each prediction task. Evaluations on simulated and real-life battery data reveal the following insights: 1) Incorporating both data types result in superior RUL prediction performance than models that only consider a single data type; 2) Using uncertainty information to weigh each prediction task results in better signal/event/RUL prediction performance than the unweighted counterparts. The proposed model is generic and can be applied to any engineering system that collects both discrete event data and continuous signal data.

3 - Decision-Aware Robust Remanufacturing Planning

JUN ZHOU, University of Houston, Houston, TX, United States, Yisha Xiang, Ming Zhao

In this study, we develop a novel decision-aware robust remanufacturing decision model to determine the optimal remanufacturing time. Due to limited data or errors in data, the estimation of transition probabilities for a product to be remanufactured is subject to large statistical errors. Robust Markov decision process is a powerful decision tool for mitigating the effects of parameter uncertainties due to estimation errors. However, conventional robust MDPs may prescribe overly conservative policies. To better control conservatism in robust remanufacturing planning, we introduce innovative decision-aware constraints to construct ambiguity sets that leverage policies from nominal models without consideration of parameter uncertainty. These decision-aware constraints are designed to reduce the search space for the worst-case transition probabilities so that the resulting policies are robust yet not overly conservative. Numerical experiments are conducted to demonstrate the utility of the decision-aware Robust remanufacturing planning model.

4 - Motion Tracking in Ultrasound Imaging via Variational Bayesian Optical Flow

Shenghao Xia, University of Arizona, Tucson, AZ, United States

Motion tracking and segmentation in carpal tunnel ultrasound imaging is a critical step in the early diagnosis the carpal tunnel syndrome. However, traditional methods heavily rely on domain experts' annotations which are expensive. Optical flow was implemented to perform motion tracking, but it lacks uncertainty quantification and fails in Rayleigh-distributed noisy ultrasound imaging. In this research, a new Bayesian optical flow method is proposed to track motion by incorporating the characteristics of ultrasound imaging. The experiments in case studies have shown the performance improvement of the proposed model over benchmarks.

TE02

Summit - 321

Smart Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Chiwoo Park, University of Washington, Seattle, WA, United States

1 - Physics-Informed Deep Learning for Steady-State Thermal Distribution Prediction Across Diverse Geometric Domains

Amirul Islam Saimon, Virginia Tech, Blacksburg, VA, United States, Raghav Gnanasambandam, Zhenyu Kong

Thermal phenomena are fundamental in many additive manufacturing (AM) processes such as laser powder bed fusion (L-PBF). Advanced understanding of thermal phenomena assists in optimizing process parameters, and subsequently, avoiding defects such as warping, residual stresses, and microstructural anomalies. Thus, thermal distribution plays a critical role in determining the microstructure, mechanical properties, and overall quality of the manufactured parts. Most of the existing thermal prediction models are restricted to single geometry, which makes the prediction for unknown geometries almost impossible. Therefore, generalizing thermal prediction model across diverse geometries is a challenging research question in the AM community. This presentation aims at introducing a deep learning-based technique for predicting steady-state thermal distribution across diverse geometries which holds promise for extending its applicability to dynamic thermal phenomena in AM processes.

2 - Modular Calibration of a Digital Twin Model for Planning-Level Decision-Making in a Semiconductor Fab's AMHS

Bonggwon Kang, Department of Industrial Engineering, Pusan National University, Busan, Korea, Republic of, Chiwoo Park, Haejoong Kim, Bosung Kim, Soondo Hong

Field practitioners in automated material handling systems (AMHSs) in semiconductor fabrication plants (fabs) widely employ digital twin models to capture highly uncertain and dynamic material handling processes for planning-level decision-making. For obtaining credible decision-making, digital twin calibration has emerged as an essential step when the models allow for all sources of uncertainty and discrepancy, e.g., code uncertainty and parametric uncertainty. In practice, field practitioners have manually adjusted tuning parameters and compensated for discrepancies in the models whenever they experience significant operational fluctuations. However, no systematic application for the AMHS digital twin calibration has been presented. This study proposes a modular Bayesian calibration for an AMHS digital twin based on the Kennedy-and-O'Hagan approach. We use surrogate models to account for the parametric uncertainty and the digital twin’s discrepancy. The experiment results indicate that the proposed approach can improve the prediction accuracy of the digital twin over a broad range of input parameters.

3 - Transformer Based Offline Printing Strategy Optimization for Large Format Additive Manufacturing

Feng Ju, Arizona State University, Tempe, AZ, United States, haoyang xie, Dylan Hoskins, Kyle Rowe

In the realm of Large Format Additive Manufacturing (LFAM), determining an effective printing strategy before actual printing involves predicting temperature behaviors and controlling layer time, which has consistently been challenging. Currently, temperature prediction for controlling layer time in LFAM is primarily conducted through offline simulations or online monitoring. However, these approaches are typically tailored to specific cases and lack generalizability. Consequently, there exists a significant gap in the development of a universal model that can leverage historical data to predict temperature across various new geometries and positions. In this paper, we propose a novel approach to optimize the printing strategies for LFAM through the development and application of a transformer-based model focused on the dynamic prediction and management of temperature profiles across the print surface. Subsequently, we input the predicted temperature into an optimization model to determine the optimal layer time. We conducted a series of experiments to validate the effectiveness of our proposed model. Using historical temperature data collected from the real printing processes, the model demonstrated a high degree of accuracy in predicting temperature profiles for new design, enabling the optimization of layer time settings far beyond the capabilities of traditional fixed-time methods. This process significantly enhances the printing strategy, thereby increasing both the efficiency of the printing process and the quality of the printed objects.

TE03

Summit - 322

Service Science Best Student Paper I

Award Session

Service Science

Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, United States

1 - Ranking Quality and User Engagement on an Online B2B Platform

Rakesh Allu, Cornell University, Ithaca, NY, United States

Online business-to-business (B2B) platforms are increasingly investing in data science teams to develop machine-learning algorithms for providing personalized rankings-based recommendations for their users. Consequently, there is growing need to develop ways to evaluate the performance of these algorithms. Two key questions faced by platforms in such evaluation are: (i) how to periodically measure the quality of rankings using real-time ranked transactions data generated by the users and, (ii) how to examine the effect of improving ranking quality on usage (i.e., the number of transactions) and browsing effort. These questions are challenging to address because of simultaneity between usage and effort and censoring in ranked transactions data. Using detailed transaction-level data from a B2B platform in India for more than 61K sellers over one year, we propose methods to measure ranking quality, develop a position-level model of a user’s decision to transact and scroll, and develop estimation approaches to overcome censoring. Our analysis reveals a position-wise dynamic that results in an increase in platform usage and a decrease in browsing effort with improvement in ranking quality. We further show that nudging a user to scroll at least up to a predefined position can enable the platform to realize higher value from improving rankings quality.

2 - AI-Powered Philanthropy: Effects on Volunteer Productivity

Vinit Tipnis, Kelley School of Business, Indiana University Bloomington, Bloomington, IN, United States

Artificial intelligence (AI) boosts employee productivity, but its effect on nonprofit volunteers, who lack monetary incentives, is unclear. This study examines AI's impact on volunteers on an online platform that recently adopted an AI tool. Using an analytical model and difference-in-differences event study design, we find AI reduces weekly contributions by 7.5% on average. The impact varies with experience: less-experienced volunteers' productivity drops, while highly experienced volunteers increase their contributions. Nonprofit managers should avoid introducing AI to inexperienced volunteers without proper education but can leverage AI for experienced volunteers and use them to advocate for AI adoption.

3 - Transparent or Not? Optimal Performance Feedback in Services

Lin Chen, INSEAD, Fontainebleau, France

In many gamified services, users receive individual performance feedback upon service completion, which shapes their perception of goal achievement (through prospect theory) and relative status (when users are ahead-seeking or behind-averse). This feedback can vary in degrees of transparency, from disclosing individual scores to revealing only ranges of scores (e.g., top 5%). How transparent should service providers be in their individual performance feedback to maximize the utility users derive from their service? We employ a Bayesian persuasion framework to determine the optimal information disclosure policy, depending on whether a goal has been specified and whether the other users' scores are communicated. Without a goal, if the other users' scores are not available, any information policy is optimal; if they are available, the provider should be fully transparent (resp., opaque) when users are ahead-seeking (resp., behind-averse). With a goal, if the other users' scores are not available, the users who have exceeded or just fallen short of the goal should only be told that they lie in that range, whereas the lowest-performing users should be told their exact scores; if the other users’ scores are available, the range of opaque disclosure may not extend to the top performers when users are ahead-seeking. Comparing the aggregate user utility across conditions, we also characterize when setting a goal or sharing the score of the other users leads to a higher aggregate utility. Our paper offers guidelines to service providers to enhance user utility by engineering the design of their relative performance feedback.

TE04

Summit - 323

Service Science Best Student/Cluster Paper II

Award Session

Service Science

Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, United States

1 - Algorithm Reliance, Fast and Slow

Clare Snyder, Michigan Ross, Ann Arbor, MI, United States

In algorithm-augmented service contexts where workers have decision authority, they face two decisions about the algorithm: whether to follow its advice, and how quickly to do so. The pressure to work quickly increases with the speed of arriving customers. In this paper, we ask: how do workers use algorithms to manage system loads? With a laboratory experiment, we find that superior algorithm quality and high system loads increase participants' willingness to use their algorithm's advice. Consequently, participants with the superior algorithm make higher-quality recommendations than those with no algorithm (participants with the inferior algorithm make slightly lower-quality recommendations than those without). However, participants do not necessarily speed up by using algorithms' advice; their throughput times only decrease compared to the no-algorithm baseline when the system load is high and algorithm quality is superior, although participants would benefit from working faster in all treatments. This happens in part because participants in the high-load, superior-algorithm treatment serve customers more quickly than participants in the other treatments, conditional on using the algorithm. Participants in the high-load, superior-algorithm treatment work especially quickly in later periods as they increasingly default to their algorithm's advice. Our findings show that algorithms can have benefits for both decision quality and speed. Quality benefits come from workers' decision to use their algorithms' advice, while speed benefits depend on workers' algorithm use and the time they spend deliberating about their algorithm use. Ultimately, algorithm quality and system load are mutually reinforcing factors that influence both service quality and especially speed.

2 - Regulating Discriminatory Pricing in the Presence of Tacit Collusion

Zongsen Yang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of

Price-setting algorithms have facilitated widespread awareness-based price discrimination, wherein firms charge high prices to customers unaware of alternative choices and low prices to those in competitive markets. This unethical behavior has increased customer complaints and prompted policymakers to enact fairness regulations in response. However, while limiting price discrimination may improve consumer welfare under genuine competition, it also affects firms' incentives to form tacit collusion, another regulatory concern arising from the proliferation of pricing algorithms. We develop an analytical model to examine the interplay between fairness regulation and tacit collusion, and discuss its impact on consumer welfare and policymaking. Firms utilize customers' product unawareness to implement price discrimination and decide whether to collude by comparing profits from collusion and deviation. We then explore the consequences of price fairness regulation on the sustainability of tacit collusion. For homogeneous products, fairness regulation can substantially weaken collusion, potentially rendering it unattainable. However, for differentiated products, strict fairness inadvertently supports collusive behavior, harming consumer welfare. In this case, mild fairness permitting moderate price differentiation can prevent market collusion and optimize welfare. Our key results are valid under various robustness checks such as incorporating Q-learning or using alternative demand models and punishment strategies. We also propose a novel randomized approach to tame fairness-induced collusion. Overall, our study emphasizes the importance of a nuanced approach to regulating discriminatory pricing in the presence of tacit collusion.

3 - Homomorphic Encrypted Revenue Management

Mojtaba Abdolmaleki, University of Michigan, Ann Arbor, MI, United States

We develop a novel homomorphic encryption-based approach to privacy preservation in a dynamic personalized pricing setting. In each period, the firm offers a personalized price to an incoming customer based on (i) this customer's observable characteristics and (ii) the firm's estimate of the demand function (obtained using the data of the historical customers with whom the firm interacted in the past). Our method enables the firm to use homomorphic encryption to encrypt the data of incoming and historical customers, then estimate the demand function and personalize prices directly based on these encrypted data without the need to decrypt them. In contrast to the previous literature, which only preserves the privacy of historical customers via adding statistical noise to their data (so-called statistics-based approach), our approach allows the firm to protect the privacy of all customers -- both incoming and historical. Our theoretical analysis further reveals that our approach i) provides perfect privacy protection (achieving 0-differential privacy) and ii) does so at no cost to the firm's expected revenue, thus achieving better revenue performance than statistics-based algorithms, but (iii) it is computationally expensive. We thus develop a hybrid approach to privacy preservation that leverages the strengths of both statistics- and encryption-based methods, achieving the required privacy protection at a comparatively lower computational cost without significant compromise on the expected revenue. We confirm our theoretical findings through a numerical example based on synthetically generated data.

TE05

Summit - 324

Service Science Best Cluster Paper I

Award Session

Service Science

1 - Operating Three-sided Marketplace: Pricing and Spatial Staffing in Food Delivery Systems

Zhe Liu, Imperial College London, London, United Kingdom, Yiwen Shen, Yanwei Sun

We study a food delivery platform's joint pricing and staffing problem under the endogenous participation of three sides—restaurants, customers, and deliverers—each with its own incentives and heterogeneous features. With deliverer pooling and nearest matching, we represent the spatial multi-server system as a state-dependent Markovian queueing model, where the service rate depends on the imbalance of all three sides due to spatial frictions. We then analyze the system's equilibrium behavior and limiting service regimes in heavy traffic, and solve the platform's revenue maximization problem in large systems. Our explicit approximate solutions circumvent the tedious exact optimization by exploiting large-system asymptotics and economic implications. We show that the platform's asymptotically optimal solutions decouple into leading terms that neglect stochastic effects, and lower-order terms that reveal a balance of server utilization and service quality by a power-of-2/3 scaling (contrasting the classic square-root scaling). The main insights gleaned from our analytical results include: both the Efficiency-Driven (ED) and Quality-Driven (QD) regimes may prevail, depending primarily on customers' delay sensitivity; the revenue-maximizing platform still generates remarkable value to all three parties as well as social welfare, thanks to deliverer pooling and spatial matching; restaurants with intermediate profit margins join the platform, reaping demand boost and logistic savings.

2 - How Forced Intervention Facilitates Long-term Algorithm Adoption

Chenshan Hu, Washington University in St. Louis, Saint Louis, MO, United States, Xinyu Cao, Dennis Zhang

While artificial intelligence (AI) technologies become increasingly powerful and useful, human workers often resist adopting algorithmic recommendations, known as algorithm aversion. This aversion undermines the algorithms' performance in practice. While numerous studies explored short-term mitigation strategies, this paper investigates whether and why forced interventions can promote long-term algorithm adoption. Data from a leading online education company reveal that sales workers underutilize a new matching algorithm. The company then conducted a field experiment where sales workers were forced to use or not use the algorithm for three weeks. Results show that this forced intervention during the experiment causally increases workers' algorithm usage one month after the experiment by 15.8 percentage points. We develop a theoretical model to derive empirical strategies for exploring the mechanisms behind this. Contrary to the traditional literature focusing on habit formation, our findings suggest that learning is a key driver for long-term algorithm adoption. Specifically, forced algorithm usage allows workers to experience the algorithm's unbiased performance and positively adjust their beliefs about it. Consequently, workers use the algorithm not only more frequently but also more on high-quality leads post-experiment. We provide empirical evidence that forced intervention can effectively improve long-term algorithm adoption. More importantly, we demonstrate that forced intervention works by enabling workers to experience an algorithm's unbiased performance and adjust their prior misinformed assumptions about its effectiveness. This suggests that firms can implement extrinsic interventions to help workers recognize the benefits of algorithms and develop unbiased beliefs about their capabilities, thus facilitating sustained algorithm usage.

3 - Workforce Configuration in Charity Settings: A Forward-looking Approach

Chao Wu, Wayne State University, Detroit, MI, United States, Mahyar Eftekhar, Joline Uichanco

Volunteers form the backbone of many charities but pose challenges due to their variability in commitment and affinity. This study builds on existing research to propose a novel volunteer scheduling model aimed at optimizing resource allocation while nurturing future donor relationships. By integrating volunteer heterogeneity and donation potential into scheduling decisions, the model mitigates staffing costs and enhances volunteer satisfaction, thereby increasing their likelihood of future donations. Strategic adoption of this integrated approach challenges traditional organizational structures, further emphasizing the importance of robust data management in advancing charitable missions.

4 - Keep the Water Flowing: The Hidden Crisis of Rural Water Management

Chengcheng Zhai, Univeristy of Notre Dame, South Bend, IN, United States, Rodney Parker, Kurt Bretthauer, Alfonso Pedraza-Martinez, Jorge Mejia

In rural sub-Saharan Africa, people rely on water points like communal handpumps for clean drinking water, but these often break down, requiring reactive repairs. Local NGOs face tough decisions on allocating limited resources to reduce water point downtime, which affects health outcomes. They could invest in gathering more functionality data, increase repair capacity, or source more reliable parts. To minimize water point downtime, we propose a comprehensive maintenance service approach, which includes preventive maintenance and reactive repairs. In collaboration with local NGOs, we conducted field research in Ethiopia and Malawi. We collected 47,240 observations on water point functionality from NGOs in Malawi, the Central African Republic, and Ethiopia. Then, we built a Markov decision process (MDP) based on real-world field practices. This approach helped us determine the optimal schedule for NGO mechanics to visit and maintain water points. We applied the model to field data from the three countries and found that comprehensive maintenance can reduce water point downtime by up to 61.9%, at a minor increase in logistics cost, if any. Our findings challenge the belief that preventive maintenance is prohibitively expensive, recommending its integration into NGO programs. We further recommend that resource-constrained NGOs prioritize improving water point reliability, increasing repair capacity, or reducing the cost of major repair before investing in extensive data collection.

TE06

Summit - 325

Service Science Best Cluster Paper II

Award Session

Service Science

Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, United States

1 - Consider or Choose? The Role and Power of Consideration Sets

Yi-Chun Akchen, University College London, London, United Kingdom, Dmitry Mitrofanov

Consideration sets play a crucial role in discrete choice modeling, where a large class of non-parametric models can be understood as a two-stage decision making process. In this process, customers first form consideration sets and then follow a choice mechanism, such as a ranking system or a decision tree, to navigate their way towards a purchase decision. Rather than delving into a more sophisticated second-stage choice mechanism, this paper takes a step back and focuses on the fundamental properties of the decision-making process based on the consideration sets. We consider a class of nonparametric choice models that is only specified by a distribution over consideration sets and has a bounded rationality interpretation. We call it the consideration set model. Intriguingly, we show that this class of choice models can be characterized by the axiom of symmetric demand cannibalization and enable full statistical identification. We further consider the model's downstream assortment planning as an application. We first provide an exact characterization of the optimal assortment and show that it is revenue-ordered on the blocks defined with respect to the consideration sets. Despite the existence of the exact structure, we prove that the optimal assortment is NP-hard to approximate. This result shows that the existence of consideration sets inevitably cause inapproximability in assortment planning, despite that the consideration set model does not involve a sophisticated second-stage mechanism. Finally, using a real-world dataset, we show that the consideration set model offers comparable prediction power to other advanced choice models.

2 - Learning the Minimal Representation of a Continuous State Space MDP from Transition Data

Mohammed Amine Bennouna, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgia Perakis, Dessislava Pachamanova

In this talk, we introduce a framework for learning the most concise (minimal) MDP model of a continuous state space dynamic system from observed transition data.
Our work is motivated by data-driven personalized patient treatment—an increasingly sought-after machine learning application. In this reinforcement learning problem, as in most high-stakes applications, interpretability is crucial: physicians cannot rely on a black-box algorithm for prescribing treatments. We introduce theoretically the problem of learning the most concise representation that preserves the system’s dynamics. In the patient treatment setting, this corresponds to identifying treatment groups based on the evolving features of patients under treatment. Surprisingly, we prove theoretically that it is statistically possible to learn the most concise representation of a dynamic system solely from observed historic sample path data. We subsequently develop an algorithm, MRL, which learns such a concise representation, thereby enhancing interpretability and tractability.

3 - Coordinate Package Delivery and On-Demand Rides: A Zoning Policy and Analysis

Junyu Cao, The University of Texas at Austin, Austin, TX, United States, Sheng Liu

Cities are facing mounting pressure from the surging demand for package delivery and on-demand ride services. Meanwhile, more drivers are empowered by platforms to provide flexible transportation and logistics services with their own vehicles. Integrating freight and passenger transportation services, termed co-modality, has been promoted as a promising means to improve vehicle utilization, driver productivity, and logistics efficiency. Motivated by these issues, we investigate the value of co-modality to the driver in the context of joint package delivery and ride-hailing from a revenue-maximizing perspective. The driver must decide when and how on-demand ride jobs should be coordinated with package deliveries. Because the general optimal coordination problem is intractable, we develop a zoning-based coordination policy that allows us to capture the synergies between these two types of jobs and characterize its revenue performance analytically. We find that the optimal policy reduces to three forms: the pure package delivery policy, the four-zone policy (limited ride opportunity), and the switching policy (flexible ride opportunity). We show that co-modality is most beneficial when the package volume is relatively low. Offering limited flexibility is optimal when the ride profit and package volume are moderate. In the limit, the zoning-based coordination policy scales similarly to the pure package delivery policy and enjoys a bounded performance ratio to the perfect coordination policy. We also complement our analysis using numerical studies based on taxi data from New York City and discuss the practical implications.

4 - Machine Learning-Augmented Optimization of Large Bilevel and Two-stage Stochastic Programs: Application to Cycling Network Design

Bo Lin, University of Toronto, North York, ON, Canada, Timothy Chan

Motivated by a cycling infrastructure planning application, we present a machine learning approach to solving bilevel programs with a large number of independent followers, which as a special case includes two-stage stochastic programming. We propose an optimization model that explicitly considers a sampled subset of followers and exploits a machine learning model to estimate the objective values of unsampled followers. Unlike existing approaches, we embed machine learning model training into the optimization problem, which allows us to employ follower features that cannot be represented using leader decisions. We prove bounds on the optimality gap of the generated leader decision as measured by the original objective that considers the full follower set. We develop follower sampling algorithms to tighten the bounds and a representation learning approach to learn follower features, which are used as inputs to our machine learning model. Through numerical studies, we show that our approach generates leader decisions of higher quality compared to baselines. Finally, we perform a real-world case study in Toronto, Canada, where we solve a cycling network design problem with over one million followers. Compared to the current practice, our approach improves a transportation metric by 19.2% and can lead to a potential cost saving of $18M.

TE07

Summit - 327

Global Optimization

Invited Session

OPT: Computational Optimization and Software

Chair: Zelda Zabinsky, University of Washington, Seattle, WA, United States

1 - Multiple model consistency for global optimization

Danielle Morey, University of Washington, Seattle, WA, United States, Giulia Pedrielli, Cherry Wakayama, Zelda Zabinsky

Complex systems are often modeled with multiple models, including simulations, Gaussian processes, and data regression. In the absence of a known ground truth model, it is important to reconcile the consistency of different models in a region of interest. In a heterogeneous space, model accuracy may depend on subregions of the space. We combine state space partitioning with model consistency measures to refine model estimation near global optima. We also incorporate ideas of statistical consistency with the concept of consistency sets in linear and convex programming to identify near-optimal regions of interest based on shared trends. While model outputs may differ, regions can be compared with regards to the global objective function via consistency and improvements across models. By evaluating the same points with each model, a point-wise comparison can be drawn to determine model consistency and can be used as an indirect measure of accuracy. As a proof of concept, we apply a multiple model consistency approach to global optimization on test problems as well as maritime and biomanufacturing applications.

2 - Scalable Global Optimization with Adaptive Surrogate Search and Subspace Decomposition

Pariyakorn Maneekul, University of Washington, Seattle, WA, United States, Giulia Pedrielli, Zelda Zabinsky

Global optimization of large-scale black-box functions is being pursued using surrogate modeling, partitioning, and adaptive sampling distributions. Recent finite-time analyses on Branching Adaptive Surrogate Search (BASSO) highlight the importance of consistently improving the incumbent objective function value. One condition of the analysis is that the implicit adaptive sampling distribution stochastically dominates a uniform sampling distribution on the level set of improving points. We explore decomposing the high-dimensional space into many low-dimensional subspaces to enable surrogate modelling within practical limits. We combine the subspace variables using an adaptive subspace distribution to create a full-dimensional point for objective function evaluation. We numerically compare the performance of several BASSO implementations with state-of-the-art methods on several benchmark objective functions with dimensions ranging from 20 up to 1,000. We also test whether the BASSO implementations satisfy the assumptions of the BASSO finite-time analysis.

3 - Algorithm-Guided Experimentation for Systematic Engineering of Perovskite Solar Cells

Donghyun Oh, Georgia Institute of Technology, Atlanta, GA, United States, Sanggyun Kim, Carlo Andrea Riccardo Perini, Juan-Pablo Correa-Baena, Nikolaos Sahinidis

Perovskite solar cells hold significant potential for renewable and sustainable energy generation. However, research into perovskite photovoltaics has been predominantly experimental and hampered by a lack of systematic approaches for optimizing key design parameters. In this work, we present a combined experimental-computational framework to systematically enhance the photovoltaic performance of these devices. By employing black-box optimization algorithms, our framework guides the sampling of the design space to identify optimal designs. We then fabricate devices based on this algorithm-guided experimental design and test them under illumination to validate and refine the optimization process. The results show notable performance enhancements, demonstrating an effective combination of mathematical optimization and experimental research.

4 - AI-Driven Laser Parameter Search (ALPS): Inverse Design of Photonic Surfaces using Optimization Techniques

Luka Grbcic, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, Juliane Muller, Wibe Albert de Jong

Photonic surfaces with tailored optical characteristics are increasingly utilized in various energy harvesting and storage systems. In this work, we introduce ALPS—a computational optimization framework specifically developed for the inverse design of photonic surfaces. ALPS uses a greedy, prediction-based exploration strategy to optimize laser fabrication parameters that are crucial for creating photonic surfaces with minimal discrepancy from user-defined target optical characteristics. We have applied ALPS to a series of inverse design benchmarks and conducted comparisons with established optimization algorithms to evaluate its performance. Our results highlight a significant advantage of ALPS: its ability to efficiently warm-start the inverse design process when adapting to new target optical characteristics.

5 - Embedding ReLU neural networks in dynamic optimization: scaling mixed-integer formulations using bounds propagation

Calvin Tsay, Imperial College London, London, United Kingdom, Philip Sosnin

Neural networks are often embedded in dynamic optimization as surrogate models of a system or as efficient controllers, e.g., approximate explicit MPC. Applications include optimization of the surrogate/controlled system itself, scheduling, or certification of (worst-case) system behavior. The latter is particularly important in safety-critical domains such as autonomous vehicle driving. Discretizing, or "unrolling", the time domain produces a large, monolithic global optimization problem. This work studies the case where all components are piecewise linear, e.g., ReLU neural networks and hybrid discrete systems, resulting in a mixed-integer linear program, or MILP. Solution as a MILP can certify global optimality but suffers from scalability issues. We discuss how these issues can be (partially) mitigated via bound-tightening techniques prior to forming the mixed-integer program, generating tighter continuous relaxations and, consequently, expediting optimisation procedures. Specifically, we investigate how bounds can be efficiently and tightly propagated through neural network layers and time steps. Using certification of neural network controllers as a case study, we evaluate several strategies for bounds tightening in terms of both computational complexity and tightness of the bounds.

TE08

Summit - 328

AI and Optimization Techniques in Forecasting and Security

Contributed Session

Chair: Sepehr Soltani, American Airlines, 4444 Cole Avenue, Dallas, TX, 75205, United States

1 - Convergence-Guaranteed Elastic Net Graphical Model Estimation for Anomaly Localization

Dzung Phan, IBM Research, Yorktown Heights, NY, United States, Tsuyoshi Ide, Matt Menickelly, Jayant Kalagnanam

Estimating dependency structures from noisy multivariate variables is crucial across various applications, with anomaly localization being particularly vital. Anomaly localization involves computing variable-wise anomaly scores by comparing a target dependency structure to a reference one. Stable and accurate estimation of dependency structures is essential for this task. Firstly, we present an L0-elastic net model designed to estimate sparse inverse covariance matrices. Subsequently, we introduce a framework for anomaly localization that integrates the L0-elastic net model with a transfer learning approach. Despite the known challenges of L0-constrained optimization, we introduce a hard thresholding line-search algorithm to efficiently solve these graphical models. Through experiments on synthetic and real-world datasets, we demonstrate that our proposed L0-based method consistently outperforms alternative methods across numerous use cases.

2 - Hierarchical Forecasting for Amazon EC2 demand

Yiming Wang, Amazon, Seattle, WA, United States, Ebrahin Nasrabadi

Amazon's Stores, Digital, and Other (SDO) is the largest user of AWS EC2 capacity. This poster introduces our in-house Hierarchical Bayesian and Hierarchical Reconciliation forecasting methods to forecast SDO EC2 demand. The demand forecasts are published at bi-monthly cadence, covering a rolling 2-year horizon, and at granularity of business group (Stores, Digital, Other), availability zone (e.g., us-east-1a), and instance type (e.g., c5.12xlarge) level. The forecasts can be used for AWS capacity planning to drive operational and financial decisions on global AWS data center infrastructure and space plan. To address the cold-start challenges in forecasting new instance generation and availability zones's demand, we developed a novel Hierarchical Bayesian time series forecasting model. In back-testing, our model was proven to reduce the forecast error measured in Weighted Absolute Percentage Error (WAPE) to 18.6% from 30.0% using Meta's open-source Prophet package for new AWS EC2 instances launched in 2023.

3 - Securing through evaluating: An autonomous cloud security assessment artifact based on large language models

Qin Su, School of Management, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Yue Fang, YING GAO

Cloud security has emerged as a pivotal concern that postpones the development of online cloud systems in multiple areas. Conventional cloud platform security assessment techniques suffer from low assessment accuracy and reliance of subjective viewpoints on security, rendering their applications in the digital era. To overcome these vulnerabilities and drawing from the design science theory, we designed a comprehensive cloud security assessment artifact enabled by large language model based autonomous system (LLMAS). The proposed artifact presents a dynamic and adaptive mechanism for continuous online platform security assessment including two distinct system architectures: the algorithm architecture and the process architecture. The algorithm architecture refers to a multi-dimensional sentiment classification generator driven by a hybrid algorithm of large language models (LLM) and domain knowledge graph, and the process architecture refers to an autonomous system security assessment process including a loop of construction-validation-evaluation steps. To validate the effectiveness of our artifact, we collected standard documents text data in the cloud ERP context and constructed a comprehensive evaluation system as an illustration. The hybrid LLM algorithm is superior to existing text analysis methods with an increase of around 27% in prediction accuracy. Our paper contributes to existing literature by promoting the applications of LLM in the cloud domain, proposing an autonomous assessment process which balances between users' heterogeneous security perception and assessment costs, and constructing a viable tool of cloud ERP security based on standard documents.

4 - A Deep Reinforcement Learning-Based Technique for Optimal Power Allocation in Multiple Access Communications

Sepehr Soltani, American Airlines, Dallas, TX, United States, Ehsan Ghafourian, Diego Martin, Milad Vahidi, Reza Salehi

Addressing the escalating demands of modern wireless networks, this study introduces a sophisticated deep Q-learning (DQL) framework to optimize power allocation (PA) in wireless multiple-access communications. Central to our approach is the employment of a model-free Deep Q-Network (DQN), which leverages the current sumrate as the reward function, aligning PA strategies with the overarching goal of maximizing network efficiency and fairness. This innovative method enhances decision-making by enabling precise sumrate predictions across diverse power settings based on specific channel state information (CSI). Our results demonstrate that the DQN with a zero discount factor not only achieves superior average sumrates but also significantly outperforms conventional methods such as fractional programming (FP) and weighted minimum mean squared error (WMMSE), especially in varied user density scenarios. The robustness of our approach is further validated through offline centralized learning, where the DQN is initially trained in simulated environments and then adeptly applied in real-time through distributed executions. Our findings demonstrate not only improved sumrate performance but also robust generalization capabilities, showcasing how advanced analytical techniques can lead to smarter resource management in next-generation wireless networks. This study underscores the potential of machine learning in enhancing the decision-making processes within complex communication systems, promoting more effective and sustainable technological advancements.

TE09

Summit - 329

Innovations in Market Design

Invited Session

Auctions and Market Design

Chair: Gerard Cachon, University of Pennsylvania, Philadelphia, PA, United States

1 - Platform Disintermediation: Information Effects and Pricing Remedies

Auyon Siddiq, University of California, Los Angeles, Los Angeles, CA, United States, Shreyas Sekar

Two-sided platforms, such as labor marketplaces for hiring freelancers, typically generate revenue by matching prospective buyers and sellers and extracting commissions from completed transactions. Disintermediation, where sellers transact offline with buyers to bypass commission fees, can undermine the viability of these platforms. Although transacting offline allows sellers to avoid commission fees, it also leaves them fully exposed to risky buyers (given the absence of the platform's payment protections) and imposes switching costs (given the absence of the platform's transaction infrastructure). In this paper, we consider interventions for addressing disintermediation, focusing on the pricing and informational levers available to the platform, where the latter refers to the accuracy of the signal sellers receive about buyers' riskiness. First, while intuition suggests platforms should counter disintermediation by lowering commission rates, in a high-information environment a platform may be better off raising them. Further, a platform may strictly benefit from sellers receiving a partially-informative buyer signal (i.e., not perfectly revealing nor concealing a buyer's riskiness), particularly when switching costs are low. Finally, while charging sellers platform-access fees can immunize the platform from disintermediation, it can fall short of the optimal revenue under commission-based pricing. We also examine the efficacy of banning sellers if caught disintermediating and extend our findings to a setting with repeated

transactions. Overall, our results shed light on how disintermediation disrupts platform operations and offers prescriptions for platforms seeking to counteract it.

2 - Optimal Auction Design with Deferred Inspection and Reward

Alexandre Belloni, Duke University, Durham, NC, United States, Saeed Alaei, Ali Makhdoumi, Azarakhsh Malekian

Consider a mechanism run by an auctioneer who can use both payment and inspection instruments to incentivize agents. The timeline of the events is as follows. Based on a pre-specified allocation rule and the reported values of agents, the auctioneer allocates the item and secures the reported values as deposits. The auctioneer then inspects the values of agents and, using a pre-specified reward rule, rewards the ones that have reported truthfully. Using techniques from convex analysis and calculus of variation, for any distribution of values, we fully characterize the optimal mechanism for a single agent. Using Border's theorem and duality, we find conditions under which our characterization extends to multiple agents. Interestingly, the optimal allocation function, unlike the classic settings without inspection, is not a thresholding strategy and instead is an increasing and continuous function of the types. We also present an implementation of our optimal auction and show that it achieves a higher revenue than auctions in classic settings without inspection. This is because the inspection enables the auctioneer to charge payments closer to agent's true value without creating incentives for them to deviate to lower types.

3 - Battery Operations in Electricity Markets: Strategic Behavior and Distortions

Jerry Anunrojwong, Columbia Business School, New York, NY, United States, Santiago Balseiro, Omar Besbes

Electric power systems are undergoing a major transformation: integrating intermittent renewable energy sources, and batteries to smooth out variations in renewable energy production. As privately-owned batteries grow from their role as marginal "price-takers" to significant players in the market, a natural question arises: How do batteries participate strategically in electricity markets, and how much do the incentives of independent batteries align with those of the system as a whole? We propose an analytically tractable model that captures salient features of the highly complex electricity market that are crucial in understanding the strategic role of batteries and their impact on system cost. We derive in closed form the resulting battery behavior and generation cost in three operating regimes: (i) no battery, (ii) centralized battery (first-best), and (ii) decentralized profit-maximizing battery (second-best). The battery can strategically withhold its capacity in three ways. First, withholding in quantity, or discharging less than socially optimal. Second, withholding in time, or discharging later than socially optimal. Third, withholding in responsiveness, or discharging less in response to smoothing real-time demand than socially optimal. We quantify each of the three forms of withholding in terms of market fundamentals. We further analyze the impact of battery competition. To illustrate our results, we calibrate our model to California and Texas markets and show that the loss from incentive misalignment can be consequential.

4 - Design of Resale Platforms

Daniela Saban, Stanford University, Stanford, CA, United States, Ilan Morgenstern, Divya Singhvi, Somya Singhvi

We study resale platforms, an emerging type of online marketplaces in developing countries. These platforms are designed for individuals (resellers) to sell products to others as opposed to buying for themselves, enabling them to supplement their income by earning a margin on the transactions they generate. One challenge these platforms face is that competition among resellers may emerge as more of them join the platform, as their social circles increasingly overlap. We investigate the mechanisms by which competition affects resellers and the role of the platform's design in shaping this relationship. We combine analytical modeling with empirical analysis. Our model captures the resellers' interactions with the platform (in searching for products to sell) and with consumers (in offering these products for purchase). We find that as competition intensifies, resellers not only lower their margins but also reduce the effort they exert in searching for products to sell. We empirically validate these equilibrium predictions using data from a major platform: as competition increases, resellers earn lower margins and reduce the number of products they view while browsing on the app, which we interpret as lower search effort. We explore two interventions that aim to benefit resellers: centralizing margin decisions (a practice adopted by several platforms) and optimizing its product ranking algorithm, offering insights to guide key design choices for resale platforms, aiming to foster sustainable online entrepreneurship in developing economies.

TE10

Summit - 330

Advances in Queueing Theory and Applications

Contributed Session

Chair: Himanshu Arha, Indian School of Business, Hyderabad, Hyderabad, India

1 - Crowdsensing for Informed Decision-Making: Exploring the Role of data contributors ratio and Congestion

Elham Heydarigharaei, York University, Guelph, ON, Canada, Hossein Abouee Mehrizi, Mehdi Nourinejad

Crowdsensing platforms collect data from individual users, reducing the need for extensive data collection infrastructure. The accuracy of the gathered data relies on the proportion of users contributing. However, due to concerns about privacy or limitations in device accessibility, only a subset of users actually contribute data. By integrating crowdsensing into service facilities, customers can share queue information, aiding others in deciding when to join the queue. Through the analysis of a single-server queue model, we examine how this shared data affects decisions regarding queue participation. Additionally, we explore the impact of the contributor ratio on social welfare. Our findings reveal that changes in social welfare relative to the contributor ratio are contingent upon the load factor—the balance between service demand and completion rate. In systems with high load factors, maintaining a contributor ratio below one maximizes social welfare, whereas in low-load factor systems, social welfare consistently improves as the contributor ratio increases. Moreover, our results suggest that limiting customer access to crowdsensing platforms diminishes social welfare.

2 - Asymptotic analysis of queueing systems with impatient customers and stochastic agent availability

Arash Asgari, University of Alberta, Edmonton, AB, Canada, Saied Samiedaluie, Amir Rastpour

Most queueing models for call centers assume that the number of available agents is known and controllable. In reality, though, the availability of human agents varies stochastically in a way that cannot be fully controlled due to a variety of factors such as post-call tasks, breaks, or meetings. We formulate a queueing system with impatient customers and stochastic agents' availability as a phase-type queueing model. We then proceed with an asymptotic analysis of such model in the Halfin-Whitt heavy-traffic regime which simultaneously addresses both important aspects of a queuing system: the quality of service delivered to customers and the system's efficiency. Guided by the asymptotic behavior, we first derive the diffusion limits characterizing a multidimensional piecewise Ornstein-Uhlenbeck process. However, deriving approximations from this complex process encounters significant challenges, primarily due to boundary issues. To address this challenge, we employ an innovative methodology to transform the multidimensional diffusion process into a tractable one-dimensional form. Solving this process allows us to propose high-quality approximations for key steady-state performance metrics, including customers' probability of delay, abandonment likelihood while waiting, and expected wait time. Our findings hold significant implications for system managers, providing actionable insights into setting staffing levels to meet desired service standards effectively.

3 - Posterior Sampling Learns the Best to Earn the Most

Mohamad Kazem Shirani Faradonbeh, Southern Methodist University, Plano, TX, United States

A classical problem in decision-making under uncertainty is that of sequentially learning to choose the best among multiple options. Each option yields a stochastic reward value that is determined by its own unknown parameter together with some time-varying contextual information. So, to sequentially learn the option of highest reward in an online fashion, one needs to try all different options to collect the so-called contextual bandit feedback data. At the same time, by experimenting options of lower rewards, we defeat our purpose of gaining the maximum possible reward over time. Accordingly, a fundamental dilemma is the trade-off between learning the best and earning the most. We present an algorithm that samples from the posterior belief about the unknown parameters and uses that to choose the seemingly best option. Our setting allows for additional uncertainties such that the context signals are observed only in a partial and noisy manner. We theoretically establish that the algorithm learns the best option fast to incur a minimal compromise in reward, and showcase that with real data experiments.

4 - Priority Queueing: Impact of Heterogenous Buffer and Service Time

Himanshu Arha, Indian School of Business, Hyderabad, Hyderabad, India, Vishwakant Malladi, Milind Sohoni

We study an M/D/1 queuing model with finite buffer capacity and two customer classes. We show that the control policy is of threshold type where a low-priority customer is held up to a threshold in anticipation of a high-priority customer. We examine how buffer capacity and service time influence the threshold. We also study the impact of visibility of future arrivals and compare the results with the no visibility case.

TE11

Summit - 331

Advances in Optimization Methods for Inventory Management

Contributed Session

Chair: Menglei Jia, N/A

1 - Multi-period data-driven inventory control: performance of deep neural network policies

xun zhang, National University of Singapore, Singapore, Singapore, Shen Guohao, Hu Xiangbin

We study finite-horizon periodic-review inventory control with dependent demands, in which the retailer has no access to the true demand distribution and makes inventory decisions based only on historical data. The demand distribution in each period is determined by some external features that have Markovian temporal dependence. It is first shown that the state-dependent base-stock policies are optimal. However, without the full knowledge of demand and covariate distributions, we must estimate these optimal order-up-to level functions from data. To this end, we look at the empirical average of the cumulative cost from each period to the terminal period and recursively minimize the empirical cost-to-go functional backward in time over the space of deep neural networks (DNN) to estimate optimal order-up-to functions. The construction of the data-driven policy is intuitive but the analysis is challenging. Due to the multi-period nature, estimation error in future periods propagates backward. This renders the existing non-parametric statistic theory inapplicable. We must leverage the special cost structure of the inventory model to establish various structural properties of the total expected cost function so that different sources of estimation error can be dissected. We establish finite-sample upper bounds of expected regret of implementing the proposed data-driven policy compared with optimal policies.
This upper bound for DNN-based policy is new in both statistics and operations research literature. It is also shown that the established upper bound matches the lower bound for arbitrary data-driven inventory policies up to some logarithm factors, and so is nearly tight.

2 - Optimal Policies for Decentralized Multi-Echelon Inventory Problems with Endogenous Markov-Modulated Demand Models

Byeong Kwon Lee, Seoul National University, Seoul, Korea, Republic of, JeongWook Lim, Kun Soo Park

We study the optimal inventory policy in a decentralized multi-echelon system, where a warehouse serves a retailer that faces independent and identically distributed customer demand. The retailer follows an (s, S) inventory policy, which causes the warehouse to experience non-stationary demand from the retailer. To effectively analyze this non-stationary retailer demand, we develop an endogenous Markov-modulated demand model. We also develop an efficient algorithm to calculate the lead-time demand distribution for the warehouse by utilizing generating function and its inversion. Based on this model, we propose a novel inventory modeling scheme incorporating impact horizon and state progression phase to address the overshoot issue arising from non-stationary retailer demand. Then, we prove the optimality of a state-dependent base-stock policy for the warehouse and propose an exact algorithm to efficiently compute the optimal base-stock levels. Our approach provides a rigorous foundation for managing inventory in a decentralized supply chain and provides a potential for significant saving in inventory costs.

3 - Optimal Production Policy Under TIME- and Price-Dependent Demand and Reliability

Sarbjit Singh, Institute of Management Technology Nagpur, Nagpur, India, Shivraj Singh, Dipti Singh

In this paper, we study an inventory system for products where demand depends on time and price. We consider the effect of reliability and a two-level trade credit. We suppose that the demand rate is a price- and time-dependent function. The objective is to determine the economic lot size and the best selling price to maximize the total profit per unit of time given a product's reliability and trade credit. We present an efficient procedure to determine the optimal solution to the inventory problem for all possible scenarios. This procedure is illustrated with several numerical examples. A sensitivity analysis of the optimal inventory policy with respect to the parameters of the demand rate function is also given. Finally, the main contributions of this paper are highlighted, and future research directions are introduced.

4 - Bandit Algorithm for Online Learning in Inventory Problems with Lead Times

JeongWook Lim, Seoul National University, Seoul, Korea, Republic of, Byeong Kwon Lee, Kun Soo Park

While data-driven approaches have been successfully utilized to many inventory problems, they are either offline learning or problem specific algorithms. In addition, while deep reinforcement learning approaches has been known to be capable of being a universal methodology for inventory problems, they do not support online learning capability. Thus motivated, we develop an efficient online inventory management algorithm that is applicable to various inventory problems universally. Our algorithm adjusts vanilla bandit algorithm with a virtual inventory cost framework, and it utilizes merits of both UCB-type algorithm and Thompson-sampling algorithm. We obtain regret upper bound of this algorithm and evaluate its performance with numerical experiments.

5 - Distributionally Robust Optimization of the Cyclic Inventory Routing Problem

Menglei Jia, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Albert Schrotenboer, Ahmadreza Marandi, Feng Chen

We study the cyclic inventory routing problem. The demand faced by retailers is stochastic, and its distribution is unknown. Such demand uncertainty is characterized by a moment-based ambiguity set. Instead of considering fixed replenishment intervals as in most cyclic-setting studies, we allow for variable intervals, which can cope with non-stationary demand and, therefore, are more relevant to practical applications but, at the same time, introduce significant complexity to the solving.

Our contributions can be summarized as follows. First, we formulate the problem with chance constraints on vehicle capacity and, under the back-ordering assumption, to minimize total transportation, holding, and backorder costs. Then, we reformulate the chance constraints into linear ones. In addition, we show that, for a given retailer and an interval, the worst-case distribution is a multi-point distribution. To solve the problem, we design a branch-and-price framework and provide a tailored labeling algorithm for the complex pricing problem. We conduct experiments on synthetic data and a case study on real-life data from our industry partner. The results show that our distributionally robust optimization setup outperforms stochastic programming that considers a single distribution. We also compare our setting with the fixed interval setting, illustrating that the variable interval setting can significantly reduce the operational cost. Our method does not require daily solving or rely on distributional assumptions, making it an efficient tool in practice.

TE12

Summit - 332

Drone Based Operations and Delivery Systems

Contributed Session

Chair: Ziping Wang, Morgan State University, Baltimore, MD, United States

1 - Characterizing Rural Resident Acceptance of Drone Delivery Services: A Large Language Model Empowered Approach

Ziping Wang, Morgan State University, Baltimore, MD, United States, Henan Zhu, Xiaozheng He

With the increasing demand for rural logistics services and the notable disparities in service provision between urban and rural areas, there arises a compelling need to explore innovative drone-based delivery solutions. This study aims at addressing the challenges hindering the adoption of drone-based delivery, due to technological and physical barriers, that consequently affect service quality for rural residents. Such disparities amplify concerns regarding delivery equity and residents' acceptance of the potential drone delivery services. Our research presents an inaugural investigation into residents' direct willingness and sentiment toward accessible drone delivery services in rural areas using a Large Language Model (LLM)-empowered machine learning framework. Leveraging the advanced LLM-driven Light Gradient-Boosting Machine (LightGBM) method, our prediction model mitigates the cognitive bias and significantly enhances the predictive accuracy of residents' acceptance categories, compared to traditional ordinal logistic regression models. Our research advances the understanding of rural residents' acceptance to adopt drone delivery services, addressing pertinent challenges within the rural logistic landscape and the evolution of the drone delivery market. Moreover, it reveals the gap between the supply of rural drone delivery services and the demand from the rural consumer base, exploring the intricate interplay between socioeconomic factors and delivery preferences. This approach fosters a comprehensive drone-based delivery ecosystem that inclusively benefits all rural residents, irrespective of their geographical location.

2 - Strategic Fleet Composition and Efficiency Prediction for Drone-Enabled Last-Mile Delivery

Bahar Dehqani Viniche, York University, Toronto, ON, Canada, Opher Baron, Oded Berman, Mehdi Nourinejad

Last-mile delivery, the final and most costly leg of the supply chain, demands innovative solutions to enhance efficiency in both urban and suburban areas. Drones offer advantages to this sector by avoiding congested routes and follow aerial pathways at higher speeds and lower cost. This study introduces models of drone-assisted delivery strategies, whereby trucks travel to designated hubs to launch and retrieve drones. This research bridges the gap of incorporating routing-based operational-level constraints, including drone flight range, coordinating truck and drone trajectories, and multi-launching- sending more than one drone from a single hub. Our analysis reveals three scenarios where delivery efficiency is restricted by truck capacity, drone range, or synchronization between trucks and drones. Nevertheless, the effectiveness of drone delivery varies across different urban settings due to network structures and socio-geographic factors. In second phase of our research, we use machine learning methods to evaluate the effectiveness of drone delivery in different urban settings. We use our interactive tool "Drone Sidekick Tool," to gather data on travel distances, times, and environmental effects. This comprehensive approach enhances our understanding of drone delivery's potential and aids in developing more efficient and adaptable drone delivery solutions.

3 - Truck and drone delivery problems under travel time uncertainty

Chung-Wei Shen, National Cheng Kung University, Tainan, Taiwan, Bo-Xuan Qiu, Chia-Yu Cheng

With the rapid development of e-commerce, the volume of goods handled by logistics companies is gradually increasing. Meanwhile, the travel time during the distribution process may change due to traffic conditions, such as road congestion leading to travel time increases. Therefore, how to complete the delivery within the predefined time window under travel time uncertainty is an issue that logistics companies are eager to solve. In recent years, the continuous advancement of automation technology has given the possibility of combining drones with trucks for delivery. To improve the robustness of the delivery network under uncertain circumstances, a systematic method is required to develop a robust solution. Few studies were focused on the distribution problem of trucks combined with drones considering travel time uncertainty. Therefore, this study intends to develop a robust optimization model for a collaborative truck-and-drone delivery under travel time uncertainty. Furthermore, an adaptive variable neighborhood search algorithm is proposed to tackle large instances. A Monte-Carlo simulation tests the robustness of the obtained solutions. The proposed models can be applied to some practical applications and can be used as a reference for logistics decision-makers to make relevant decisions.

4 - On the Design of Electric Charging Infrastructure for Drones

Fateme Hafizi, Illinois Institute of Technology, Chicago, IL, United States, Mohammad Miralinaghi

Freight transportation stands as a critical contributor to urban traffic congestion and air pollution, especially with the growth of electronic commerce. To address this challenge, both government and logistics corporations are actively exploring sustainable alternatives to traditional delivery systems. Among these alternatives, electric, biofuel, and hydrogen-fueled transportation have garnered attention for their potential to reduce emissions and alleviate urban congestion. Drone innovations offer promising solutions because of their inherent advantages, including lower costs and increased accessibility, especially in areas with challenging terrain. However, one significant limitation of drones is their limited flying range, which presents a significant barrier to their widespread adoption for urban parcel delivery. To overcome this constraint, the strategic placement of charging points has emerged as a viable solution to enhance the operational capabilities of drones. This study presents a theoretical framework that optimizes the location of drone charging stations in urban areas. By leveraging advanced optimization models, this framework facilitates the effective allocation of charging locations, thereby addressing the challenge of managing delivery demands. The findings of this study offer valuable insights for metropolitan authorities, logistics companies, and urban planners seeking to mitigate congestion and improve air quality in urban settings. By embracing innovative solutions like optimized charging station placement, cities can pave the way for a more sustainable and efficient future in urban logistics.

TE13

Summit - 333

Advanced Inventory and Supply Chain Management

Contributed Session

Chair: Ishtiak Sikder, North Carolina State University, Raleigh, NC, United States

1 - An always feasible multi-period sourcing and mixing model with inventory considerations

Brian Schaefer, SimpleRose, St. Louis, MO, United States

We developed a linear, mixed integer mathematical optimization model expected to yield savings in freight cost of 6%, or $70,000 annually, for a food company. We formulated the model such that it will always return a feasible solution to the customer by using soft constraints, which are assigned a penalty factor in the objective function. The user can vary the penalty factors to prioritize certain outcomes that they desire.

Inputs to the model include a 12-month forecast for raw materials, supplier characteristics, and capacity and safety stock constraints at the production plants. Binary decision variables are used to determine which production facility each supplier should ship their raw materials. Continuous decision variables are used to determine which mix of products should be used each month as well as to track the deviations from the target values for the soft constraints. The model is constructed using Python and solutions within 1% of optimal can be found in under five minutes for 25 suppliers, 4 production facilities, 2 product types, and 12 time periods.

2 - Risk Aversion in a Data-Driven Multi-Period Inventory Control Problem

Xianghua Jiang, National university of Singapore, Singapore, Singapore, Zhisheng Ye, Loon Ching Tang, Xun Zhang

We study the multi-period risk-averse inventory control problem in a data-driven setting. In this problem, a risk-averse retailer makes periodic decisions on inventory levels based only on historical demand observations without full knowledge of the demand distribution. We adopt the popular nested formulation for risk-averse programs to formulate this multi-period problem and its data-driven counterpart under a coherent risk measure. Our objective is to study the sample complexity bound such that with high probability, the data-driven policy is near-optimal, i.e., the relative error of risk under the data-driven policy compared with the optimal risk is arbitrarily small. Analysis of this problem is inherently challenging, because the multi-period nature requires solving the risk-averse program and its data-driven version recursively backward in time, while the (empirical) risk-to-go functions in this process do not have closed-form derivatives for most risk measures, which renders existing first-order methods for the risk-neutral newsvendor model invalid. In this study, we develop a zero-order framework to establish the complexity bound on sample sizes to guarantee near-optimality of the data-driven policy with given accuracy levels. Instead of using first-order derivative information on the risk-to-go function, our analysis directly examines the class of functions that underpin each cumulative risk function and derives maximum inequalities for this functional class by computing the covering numbers. Finite-sample complexity bounds are then used to establish asymptotic properties of the estimated risk, including consistency and convergence rate. To speed computation, we propose an approximation scheme to obtain a modified data-driven inventory policy.

3 - A Clustering-based Matheuristic Approach for A Two-stage Production and Inventory Planning Problem with Family-based Setups

Jiawen Cui, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, 110819, Shenyang, China, People's Republic of, Lixin Tang, Gongshu Wang

This study investigates a two-stage production and inventory planning problem with family-based setups. In the first stage, raw materials are processed into semi-finished products, which are then transformed into final products during the second stage. Products with similar characteristics are grouped into families during processing, allowing them to share the family-based setup. The study considers different grouping rules for each stage. A linear programming model is established to address this problem. In large-scale instances, the problem becomes computationally challenging to solve due to the combination of different family groupings across stages and delivery time windows of orders. To tackle this complexity, we propose a clustering-based matheuristic approach. By grouping orders with similar delivery time windows into clusters, we can significantly reduce the total number of orders and get a relaxation of the problem. This reduced-size problem is easier to solve and a lower bound can be found. To refine the solution, we apply a matheuristic algorithm to adjust the incumbent solution values, thereby approximating the solution to the original problem. We also apply the well-known relax-and-fix heuristic to quickly obtain feasible solutions on large instances. Extensive computational experiments are conducted to evaluate the effectiveness of the model and the approach. Experimental results show that the combination of this clustering-based approach and relax-and-fix is highly effective, consistently yielding promising solutions.

4 - Optimal Production and Inventory Decisions for End of Life Products

Ishtiak Sikder, North Carolina State University, Raleigh, NC, United States, Russell King

The typical life cycle of a given product can usually be divided into three distinct phases based on market demand. The initial ”ramp-up” phase would see an upward increase in product demand upon market launch. A steady-state phase then ensues which covers the majority of the product life when demand is stable. The End-of-Life (EOL) phase comes last when demand would start to decline and eventually reach a value of zero with said product line being discontinued.

Throughout all three phases manufacturers run production cycles to produce lot sizes of their product and meet current market demand. This demand is expected to be sporadic during the EOL phase and as such running production cycles throughout the entire phase can drastically increase production and inventory costs. In this paper we propose an optimization model to find a time limit during the EOL phase by when it would be economically optimal to discontinue a given product line. An opportunity cost for the rest of the EOL phase is assumed under the convex cost function being minimized. A linear decrease in demand is assumed with extensions considered for non-linear demand trends, namely convex and concave.

TE14

Summit - 334

Advancements in Grocery Retail Operations

Contributed Session

Chair: Sunil Chopra, Northwestern University

1 - A Multi-Echelon Framework for Ordering Prepared Items in Grocery

Andy Chen, Afresh Technologies, San Francisco, CA, United States

Grocery stores often transform ingredients, such as whole fruit or meat subprimals, into prepared items, such as fruit salads or cut steaks. In recent years, the popularity of these items have continued to grow. Today, in-store prepared items can comprise up to 30% of total sales depending on the department. However, since data representing these transformations is limited, ordering their ingredients adds complexity to the ordering process.

We present a multi-echelon framework for representing in-store prepared items in a full ML-OR ordering pipeline. Providing optimal ordering decisions for single items requires understanding factors such as item inventories, floor capacities, and shelf lives. These concepts do not immediately generalize to a multi-echelon framework; for example, items may change in perishability once they are cut. We discuss how these concepts can be adapted and reimagined in the multi-echelon context to serve optimal ordering of ingredients. We then discuss empirical results from implementing this framework at scale, impacting 2% of all produce ordered in the United States.

2 - Sales Elasticity and Merchandizing Standards In Grocery

Randy Jia, Afresh Technologies, San Francisco, CA, United States, Emily Meigs

The prevailing assumption in the grocery business is that shelf fullness is essential - that fuller shelves lead to more sales. Furthermore, grocers also assume that the prominence (type, location, etc.) of the display also influences how well a particular item sells. For example, “promotional” items may be placed in a special “front of store” display that is expected to drive additional sales. While the assumptions surrounding these merchandising decisions are intuitive and completely reasonable, unfortunately, they may lead to over-ordering and over-stocking if applied too liberally. Not only does this lead to food waste, a growing global concern as upwards of 30% of fresh produce is being thrown away in U.S. grocery stores, but this cuts into profits and harms the grocer’s bottom line. In this work, we partner with our U.S. grocery store customers, and use actual sales and merchandising data to study the empirical effect of merchandising standards on sales and food waste. Of particular interest is measuring the elasticity of demand with respect to factors such as display size and shelf fullness. We leverage our data-driven insights to determine and recommend best practices on merchandising standards for our customers, reducing food waste and increasing their financial performance.

3 - A Cut Above: Representing and Recommending on Complex Cutting in Grocery Stores

Mark Velednitsky, Afresh Technologies, Bothell, WA, United States

Meat departments in grocery stores undertake significant in-store preparation, involving the transformation of large meat chunks (sub-primals) into retail cuts for customer sale. However, this process presents numerous challenges for effective inventory management.

Foremost among these challenges is the sheer variety of possible "cutting plans" — different methods for dividing a sub-primal into retail cuts. Even when executing identical cuts, different meat managers may yield varied results based on their individual skill levels. Moreover, the act of cutting the sub-primal transforms it from a storable entity into a set of retail cuts with limited shelf life. Conversely, labor constraints necessitate careful consideration, as only a finite amount of cutting can be accomplished each day.

In this presentation, we outline a comprehensive solution for managing the complexities of in-store meat cutting. Beginning with the development of robust data models to represent diverse cutting plans, our approach culminates in a system that provides actionable production planning recommendations. Ultimately, our system not only maintains sales rates but also achieves a remarkable waste reduction of over 10%.

4 - A “Smart Predict, then Optimize” application to Multiproduct Pricing Problem

Matias Marcos, Universidad Adolfo Ibañez, Concón, Chile, Luis Aburto

One of the primary activities in the retail industry involves obtaining the demand for various products and calculating daily prices, which is known as Multiproduct Pricing Optimization Problem. The significance of this issue lies in estimating both own and cross-price elasticities. However, due to incomplete information regarding external and internal factors affecting demand, biases and endogeneity issues arise. Consequently, under or overestimated elasticities lead to erroneous pricing decisions, significantly impacting retailer sales.

To address this challenge in the predictive phase, various approaches in the field of Sequential Learning Optimization (SLO) have been explored, including different types of regressions and machine learning models. Following the training of predictive models, the estimated demand is utilized as an input coefficient in the objective function of the pricing optimization problem. Despite the potential for accurate predictions, price decisions often tend to be overestimated compared to values observed in competitive price markets. Therefore, this research focuses on a novel and contemporary paradigm known as Integrated Learning Optimization (ILO) or End-to-end Learning. In this approach, predictive models are trained to obtain better decisions. A significant contribution in this realm is "Smart Predict, the Optimize" (SPO), which emphasizes minimizing a loss function oriented towards optimization.

Our methodology aligns with the SPO framework, and its performance is compared with other two-stage approaches. Leveraging a database containing prices and demand for different orange juices, our preliminary results indicate a robust result in sales between 4.18% and 9.57% respect to overestimated sequential models.

Key words: Pricing, Retail, Integrated Learning Optimization.

5 - Fast-Food Stores with a Drive-Through Recovered Post-Pandemic; Stores Without DiD Not

Sunil Chopra, Northwestern University, Evanston, IL, United States, Partha Mishra, Ioannis Stamatopoulos

We document a profound, permanent change in the US fast-food consumer demand after the COVID-19 pandemic. In short, after a big pandemic slump, visits to drive-through stores almost recovered to pre-pandemic levels, but visits to non-drive-through stores stayed permanently suppressed. We use store-visit data between 2018 and 2022 from McDonald's, Starbucks, and Dunkin' Donuts, accounting for about 10% of all US fast-food stores. Comparing December 2019 to December 2022, average monthly visits to drive-through stores changed by a moderate -4.43% (CI: [-6.46%, -2.40%]). Meanwhile, non-drive-through-store visits changed by a massive -48.14% post-pandemic (CI: [-52.33%, -44.17%]). Consistent with drive-through usage being the operating mechanism, this differential recovery pattern was driven by a 9.36% increase in short-duration visits at drive-through stores (CI: [7.15%, 11.55%]). These results' magnitudes survive a more thorough difference-in-difference analysis as well as matching on store observables, and their statistical significance survives placebo inference. For perspective, the effects we document are analogous to a migration of 25% of all Starbucks' customers and 50% of its total revenue from non-drive-through stores to drive-through stores.

TE15

Summit - 335

Frontiers in Data-driven Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Tu Ni, Harvard Business School, Boston, United States

Co-Chair: Tu Ni, Harvard, Boston, MA, United States

1 - Mitigating the Spiral Down Effect: Online Learning Under Branded Fare Structure

Lan Wu, National University of Singapore, Singapore, Singapore, Changchun Liu, Chung-piaw Teo, Xun Zhang

In the dynamic landscape of airline revenue management, accurately predicting customer behavior and adjusting ticket-selling strategies in response to fluctuating demand is critical. Traditional models often overlook intricate customer buy-down patterns, resulting in a detrimental downward spiral of revenue. Furthermore, the assumption prevalent in much of the existing literature—that arrival rates and purchasing probabilities are known a priori—is rarely practical. This study combines the adjusted fare concept with the Bayes Selector algorithm to create an innovative approach for optimizing airline ticket-selling strategies under a mixed fare structure. By dynamically adjusting fare structures based on purchase probabilities and expected profits, and utilizing the probabilistic estimates and adaptability of the Bayes Selector algorithm, this integrated methodology enables airlines to dynamically refine their seat allocation strategy for maximum profitability. Our findings show that constant regret is attainable for the problem under the independent differentiated product demand setting. More interestingly, we establish a logarithmic regret bound in the case of a mixed fare structure, where tickets are priced based on the inclusion or exclusion of additional services, such as checked baggage, seat selection, meals, and flight changes or cancellations. Furthermore, we illustrate the practical application of our approach through extensive simulation and a case study under a real-world airline revenue management scenario. This study concludes that by learning and adapting to the nuances of customer behavior, airlines can significantly enhance their revenue management capabilities, leading to more accurate demand estimation and seat allocation strategies in alignment with the existing environment.

2 - Reinforcement Learning for Intensity Control: An Application to Choice-Based Network Revenue Management

HUILING MENG, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Ningyuan Chen, Xuefeng Gao

Intensity control is a class of dynamic optimization problems featuring a continuous-time horizon and a discrete state space. Many well-known problems in Operations Research including queueing control and dynamic pricing/assortment in revenue management fall into this category. To apply reinforcement learning (RL) as a computational framework to intensity control, the major obstacle is the continuous-time horizon, as RL is mostly developed for discrete-time MDPs. Naïve uniform discretization of the time horizon is typically used, without a clear understanding of the stability, convergence or computational trade-off of such discretization. In this study, we propose to use the intrinsic discretization provided by the realized jump process and adapt RL algorithms to the continuous-time intensity control problems. By focusing on the choice-based network revenue management as a case study, we theoretically analyze the framework and develop the corresponding policy evaluation (Monte Carlo and TD), policy gradient, and Actor-Critic algorithms. Our approach is computationally efficient and does not have discretization errors compared to naïve discretization when the function approximations are chosen properly. In the numerical experiments, we show the strength of our approach compared to a number of benchmarks in the literature, including choice-based deterministic linear programming and approximate dynamic programming.

3 - From Data to Donations: Optimal Fundraising Campaigns for Non-Profit Organizations

Zhengchao Wang, Imperial College Business School, London, United Kingdom, Heikki Peura, Wolfram Wiesemann

Non-profit organizations play a crucial role in mitigating social problems around the world, but often rely on raising funds from donors through costly campaigns for their financial viability. We collaborate with a major international charity to develop and test a data-driven approach to improve the targeting of their campaigns to boost net revenue. We develop novel dynamic clustering approaches, based on the amount of historical donation data available, to learn donors' preferences among campaigns for different causes, estimate donation probabilities, and inform customized campaign strategies. Besides showing the value off this approach theoretically, we demonstrate its effectiveness empirically using both synthetic and real data from the partner organization. Our approach significantly increases net revenues by improving the speed and accuracy of estimating donor campaign preferences, thus demonstrating the potential for data-driven approaches to improve the success of non-profit fundraising campaigns.

4 - Order Stacking in On-Demand Delivery Platforms

James Scott, Duke University, Durham, NC, United States, Bora Keskin, Robert Swinney

Food delivery platforms can "stack" orders together and assign them to a single driver to pick them up and deliver them sequentially. Using a theoretical model in which each entity participates only when it is rational for them to do so, we investigate the impact of order stacking on customers, drivers, and the platform.

5 - Data-Driven Knowledge Transfer in Batch Q* Learning

Wenbo Jing, New York University, New York, NY, United States, Elynn Chen, Xi Chen

In data-driven decision-making in marketing, healthcare, and education, it is desirable to utilize a large amount of data from existing ventures to navigate high-dimensional feature spaces and address data scarcity in new ventures. We explore knowledge transfer in dynamic decision-making by concentrating on batch stationary environments and formally defining task discrepancies through the lens of Markov decision processes (MDPs).

We propose a framework of Transferred Fitted Q-Iteration algorithm with general function approximation, enabling the direct estimation of the optimal action-state function Q* using both target and source data. We establish the relationship between statistical performance and MDP task discrepancy under sieve approximation, shedding light on the impact of source and target sample sizes and task discrepancy on the effectiveness of knowledge transfer. We show that the final learning error of the Q* function is significantly improved from the single task rate both theoretically and empirically.

TE16

Summit - 336

Electricity Market Pricing focusing on the impact of uncertainties and competition on day-ahead market pricing efficiency

Invited Session

Revenue Management and Pricing

Chair: Subhojit Biswas, Texas A&M University, College Station, TX

1 - Analyzing Day-Ahead vs. Real-TIME Price Mismatch and Mitigating Risks

Subhojit Biswas, Texas A&M University, College Station, TX, United States, Bahar Cavdar, Alfredo Garcia, Joseph Geunes

Day-ahead markets for electrical power allow participants to reserve capacity for the following day, while real-time markets provide immediate pricing based on real-time supply and demand conditions. Renewable energy sources such as wind and solar have injected substantial uncertainty and complexity into energy markets because of their dependence on unpredictable weather and environmental factors. This dynamic has given rise to apparent arbitrage opportunities resulting from differences between day-ahead and real-time prices. Market participants are increasingly seeking to harness these opportunities to optimize energy storage and trading strategies, but the inherent risks associated with these strategies are not well understood. We analyze prior industry data and characterize the degree of mismatch between day-ahead and real-time prices. We then propose a model that uses risk measures to explain price differences and quantify risk-related costs. By investigating the dynamics of day-ahead and real-time prices, we propose strategies that may help market participants navigate the inherent uncertainties stemming from renewable energy variations and the associated system disruptions.

2 - Electricity Market-Clearing with Extreme Events

Tomas Tapia, Johns Hopkins University, Baltimore, MD, United States, Yury Dvorkin

Extreme events threaten power network operations, causing beyond-design failures, abrupt variations, and supply interruptions. Existing market designs fail to internalize and provide a complete risk assessment for extreme and rare events. Efficiently maintaining the reliability of renewable-dominant power systems during extreme events requires co-optimizing system resources while differentiating between large/rare and small/frequent deviations. To address this research gap, we propose a new approach using novel developments in Large Deviation Theory (LDT) integrated into chance-constrained OPF models (LDT-CC OPF). This approach aims to efficiently manage the uncertainties associated with extreme events by offering a reliable and computationally tractable solution, capable of producing a competitive equilibrium. Moreover, to improve preparedness for extreme events, we integrate extreme event statistics endogenously via including LDT-CC to model and price extreme events and use weighted chance constraints (WCC) to reduce its conservatism.

3 - Abscores, Managing Risk and Uncertainty in Electricity Systems Using Banking Scoring and Rating Methodologies

Alberto Lamadrid, Lehigh University, Bethlehem, PA, United States

We present a usable multi-layered framework that facilitates the implementation of data-enabled recourse actions, instead of worst-case risk management, and handles prospective decisions about the system dispatch and planning. The ABSCORES platform establishes an Electric Assets Risk Bureau, with different scores customized according to the application required. We leverage scoring and ratings methodologies from banking and financial institutions alongside current optimization methods in dispatching power systems to help system operators and electricity markets schedule resources. This approach is based on the observation that there are major discrepancies between the power scheduled by a system operator and the actual power generated/consumed. These discrepancies—exacerbated by unplanned contingencies (e.g., natural disasters)—are caused by multiple factors, including the different financial, environmental, and risk preferences of power producers, consumers, and aggregators. This presentation is made on behalf of the ABSCORES team at Lehigh, MIT, Argonne and Livermore National Laboratorie

4 - The role of forward markets in decarbonizing electricity markets

Dongwei Zhao, Argonne National Laboratory, Cambridge, MA, United States

This work studies the role of forward markets for decarbonizing electricity markets. We model and analyze the strategic behaviour of energy suppliers, inlcuding renewable energy and enegy storage operators, in both the forward market and the spot market. We build up a two-stage game-theoretic model for such strategic behaviors and compute the market equilibirum. We demonsrate how the existence of forward markets can improve market efficiency and reduce market prices. Besides, we explore how the uncertainty of renewables and the flexbility of storage can impact the marekt equilibrium.

TE17

Summit - 337

Inventory Replenishment in Marketplaces

Invited Session

Revenue Management and Pricing

Chair: Deniz Akturk, Washington University in St. Louis, 1 Snow Way Dr, St. Louis, MO, 63130, United States

1 - Managing Inventory in a Network: Performance Bounds for Simple Policies

Deniz Akturk, Washington University in St. Louis, St. Louis, MO, United States

We consider the joint optimization of rebalancing/sourcing inventory on a graph. We focus on the lost-sales setting with customer induced relocations. Through a coupling analysis, we provide worst-case performance bounds, with tight instances, for policies commonly used in practice. We provide further insights on the performance of these policies and discuss cost regimes where they are effective.

2 - Learning and Control of Inventories and Allocations in Multi-Warehouse Multi-Store Systems

Recep Yusuf Bekci, University of Waterloo, Montreal, ON, Canada, Mehmet Gumus, Sentao Miao, Yining Wang

With ever changing market dynamics and growing supply chain networks (i.e., warehouses and stores) for large retailers, it is crucial to study the inventory replenishment and allocation decision with unknown demand under a multi-warehouse multi-store (MWMS) network. This paper proposes a novel primal-dual learning algorithm, to tackle this challenge when demand is censored. Regret analysis shows that our algorithm achieves a good regret bound.

3 - Asymptotic Optimality of Simple Replenishment Policies for A Lost-Sales Inventory System with Delivery Lead TIME and Purchase Returns

Jinzhi Bu, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Huanan Zhang, Stefanus Jasin

We study a multi-period inventory control problem with a fixed replenishment lead time and stochastic returns. Each unit of sales can be returned within a fixed return window after the purchase time, and the returned unit can be immediately used to fulfill new demands. Unmet demand in each period is lost. The objective is to find a replenishment policy that minimizes the long-run average cost. Due to the presence of stochastic returns, even the optimal policy for the system with zero lead time is already complex since we need to keep track of the volume of unreturned sales in past periods. Naturally, computing the optimal cost is numerically intractable. In this paper, we focus on analyzing the performance of two simple replenishment policies: (1) the “base-stock policy”, which orders to maintain a constant inventory level or position in each period, and (2) the “myopic policy”, which orders to minimize the expected cost for the period when the order arrives. For both systems with zero and a positive lead time, we prove that when the so-called “unreturned lead-time demand” satisfies a mild condition, both base-stock and myopic policies are asymptotically optimal as the unit lost-sales penalty cost goes to infinity. This suggests their potential use for practical implementation, especially for retail applications where the target service level is known to be very high. Our numerical results further demonstrate that, by appropriately incorporating the return processing cost and return probabilities, the empirical performances of both base-stock and myopic policies can be improved.

4 - Exact Characterization of the Jointly Optimal Restocking and Auditing Policy in Inventory Systems with Record Inaccuracy

Naveed Chehrazi, Olin Business School, SAINT LOUIS, MO, United States

We present a continuous-time stochastic model of an inventory system with record inaccuracy. In this formulation, demand is modeled by a point process and is observable only when it leads to sales. In addition to demand that can reduce the stock, an unobservable stochastic loss process can also reduce the stock. The retailer’s goal is to identify the restocking and auditing policy that minimizes the expected discounted cost of carrying a product over an infinite horizon. We analytically characterize the optimal restocking and jointly optimal auditing policy. We prove that the optimal restocking policy is a threshold policy. Our proof of this result is based on a coupling argument that is valid for any demand and loss model. Unlike the optimal restocking policy, the jointly optimal auditing policy is not of threshold type. We show that a complete proof of this statement cannot be obtained by solely resorting to the first-order stochastic dominance property of the Bayesian shelf stock distribution induced by the demand and loss process. Instead, our characterization of the jointly optimal auditing policy is based on proving that the dynamics of the shelf stock distribution constitute a (strictly) sign-regular kernel. To our knowledge, this is the first paper that characterizes the optimal policy of a complex control problem by establishing sign regularity of its underlying Markovian dynamics.

TE18

Summit - 338

Learning and Pricing in Operations Management

Invited Session

Revenue Management and Pricing

Chair: Amy Guo, UC Berkeley, Berkeley, CA, United States

Co-Chair: Zuo-Jun Shen, JD.com & University of Hong Kong, Moraga, CA, United States

1 - Learning and Collusion in Multi-Unit Carbon Auctions

Negin Golrezaei, Massachusetts Institute of Technology, Cambridge, MA, United States, Simina Brânzei, Mahsa Derakhshan, Yanjun Han

We consider repeated multi-unit auctions with uniform pricing, which are widely used in practice for allocating goods such as carbon licenses. In each round, K identical units of a good are sold to a group of buyers that have valuations with diminishing marginal returns. The buyers submit bids for the units, and then a price p is set per unit so that all the units are sold. We consider two variants of the auction, where the price is set to the K-th highest bid and (K+1)-st highest bid, respectively.

We analyze the properties of this auction in both the offline and online settings. In the offline setting, we consider the problem that one player i is facing: given access to a data set that contains the bids submitted by competitors in past auctions, find a bid vector that maximizes player i's cumulative utility on the data set. We design a polynomial time algorithm for this problem, by showing it is equivalent to finding a maximum-weight path on a carefully constructed directed acyclic graph.

Leveraging our offline algorithm, we design efficient online learning algorithms for bidding. The algorithms have sublinear regret, under both full information and bandit feedback structures. We complement our online learning algorithms with regret lower bounds.

Finally, we analyze the quality of the equilibria in the worst case through the lens of the core solution concept in the game among the bidders.

2 - PASTA: A General Framework for Offline Assortment Optimization

Juncheng Dong, Duke University, Durham, NC, United States, Weibin Mo, Zhengling Qi, Cong Shi, Ethan Fang, Vahid Tarokh

We investigate a class of assortment optimization problems in an offline, data-driven setting. The combinatorial nature of these problems often results in insufficient data coverage, creating challenges in designing provably effective solutions. To address this, we introduce the Pessimistic Assortment Optimization (PASTA) framework. PASTA, grounded in the principle of pessimism, can accurately identify the optimal assortment, requiring only that the offline data includes the optimal assortment under general conditions. We establish the first finite-sample regret bounds for offline assortment optimization across widely used models, including the multinomial logit (MNL), nested logit, and mixed logit models. Additionally, we derive a regret lower bound for the MNL model, proving that the framework is information-theoretically optimal. Numerical studies show that PASTA consistently outperforms existing baseline methods.

3 - Last-Iterate Convergence in No-Regret Learning: Games with Reference Effects Under Logit Demand

Mengzi Amy Guo, University of California, Berkeley, Berkeley, CA, United States, Donghao Ying, Javad Lavaei, Zuo-Jun Shen

This work is dedicated to the algorithm design in an oligopoly price competition, with the primary goal of examining long-run market behavior. We consider n firms engaging in a multi-period price competition within a partial information setting---each firm can only access its first-order feedback and lacks information about competitors---under the influence of reference effects. Consumers assess their willingness to pay by comparing the current price against the memory-based reference price, and their choices follow MNL model. We use the notion of stationary Nash equilibrium (SNE), defined as the fixed point of the equilibrium pricing policy, to simultaneously capture the long-run equilibrium and stability. With loss-neutral reference effects, we propose the online projected gradient ascent (OPGA) algorithm, where each firm adjusts the price using the first-order derivatives of its log-revenues. Despite the absence of typical properties required for the convergence of online games, such as strong monotonicity and variational stability, we demonstrate that under diminishing step-sizes, the price and reference price paths generated by the OPGA attain last-iterate convergence to the unique SNE, and thereby guarantee the no-regret learning. Moreover, with appropriate step-sizes, we prove that this algorithm exhibits a convergence rate of \tilde{O}(1/t^2) and achieves a constant dynamic regret. The asymmetry nature of reference effects motivates the exploration beyond loss-neutrality. When loss-averse reference effects are introduced, we propose the conservative-OPGA (C-OPGA) algorithm to handle the non-smooth revenue functions and show that the price and reference price achieve last-iterate convergence to the set of SNEs with the rate of O(1/\sqrt{t}).

4 - Algorithmic Decision-Making Safeguarded by Human Knowledge

Wenhao LI, Shanghai University of Finance and Economic, Shanghai, China, People's Republic of, Ningyuan Chen, Ming Hu

Commercial AI solutions provide analysts and managers with data-driven business intelligence for a wide range of decisions, such as demand forecasting and pricing. However, human analysts may have their own insights and experiences about the decision-making that is at odds with the algorithmic recommendation. In view of such a conflict, we study the problems in which humans and AI interact in the decision-making process and characterize when human knowledge adds value to AI decision-making. In this paper, we provide a general analytical framework to study the augmentation of algorithmic decisions with human knowledge: the analyst uses the knowledge to set a guardrail by which the algorithmic decision is clipped if the algorithmic output is out of bound and seems unreasonable. We show that when the algorithmic decision is asymptotically optimal with large data, the non-data-driven human guardrail usually provides no benefit. However, we point out three common pitfalls of the algorithmic decision: (1) lack of domain knowledge, such as the market competition, (2) model misspecification, and (3) data contamination. In these cases, even with sufficient data, the augmentation from human knowledge can still improve the performance of the algorithmic decision. We propose a model to capture a practical and pervasive type of human-AI interaction in the decision-making process. We derive insights on when should the human analyst follow the algorithmic recommendation. We conclude that even in the era of big data, human knowledge can still play an important role and contribute to decision-making.

TE19

Summit - 339

On Recommendation Systems

Contributed Session

Chair: Ernan Haruvy, McGill University, Montreal, QC, Canada

1 - Music festivals and smart tourism: big data analysis from communication operators

Mengyin Li, China Conservatory of Music, Beijing, China, People's Republic of, Nan Chen, Linyan Liu, Ying Li

Festivals and special events can attract people to gather in a short time, promote the development of local tourism. It is known to have a great impact on the local community, society, economy, and culture. It also brings challenges to destination management and operation. Communication operators' big data information can real-timely, dynamically, completely and systematically record population flow routs. This paper analyzes the characteristics and population migration of music festival visitors through big data information of communication operators.It also discusses how these quantitative findings fit into tourism studies and research on public value creation by cultural industries. With festivals occupying a central position in the music industry, investigations into the data they generate hold opportunities for researchers to better understand industry dynamics and cultural impact, and for organizers, policymakers, and industry actors to make more informed, data-driven decisions. We hope our proposed methodological approach to festival data paves way for more comprehensive music festival studies and large-scale cultural event analytics in general.

2 - Graph Machine Learning to Estimate the Value of the Lower Bound of Orienteering Problems

Krystal Coleman, SUNY University at Buffalo, Buffalo, NY, United States, Rajan Batta, Prashant Sankaran, Moises Sudit

Orienteering, an international competitive sport, mirrors real-life challenges where agents must select a permutation of locations to visit to maximize their reward. The mathematical formulation of this, the Orienteering Problem (OP), has real-world applications in maintenance and tourist excursion planning. Optimal solutions can be found with a branch-and-cut algorithm. This is computationally expensive. In this study we exploit the graph properties of intermediate problems, using machine learning to predict the value of the lower bound of the OP. These predictions are fed to optimizing solvers to accelerate the solution process. Furthermore, when the OP instance is more complex or urgency is required, or both, heuristics such as ant colony optimization and local search can be employed to generate solutions. In this work our predicted lower bounds empower decision makers by providing a standard by which to evaluate heuristic solutions to the OP.

3 - Optimal Design of Bad Recommendations

Ernan Haruvy, McGill University, Montreal, QC, Canada, Yefim Roth

Using the insight that decision makers retrieve historical rank information more easily than historical payoffs, we argue and demonstrate that this behavioral pattern leaves decision makers exploitable by recommendation systems which recommend inferior actions, as long as the recommendations are correct most of the time. In a newsvendor setting, we show that experimental participants learn to follow recommendation systems which recommend the higher payoff action most of the time, even when the recommendation is on average disadvantageous.

TE20

Summit - 340

Strategy in Mechanism Design

Invited Session

Decision Analysis Society

Chair: Rojin Rezvan, University of Texas at Austin, Austin, TX, United States

1 - Sequential Strategic Screening

Ali Vakilian, Toyota Technological Institute at Chicago (TTIC), Chicago, IL, United States

We initiate the study of strategic behavior in screening processes with multiple classifiers.
We focus on two contrasting settings: a ``conjunctive" setting in which an individual must satisfy all classifiers simultaneously, and a sequential setting in which an individual to succeed must satisfy classifiers one at a time. In other words, we introduce the combination of strategic classification with screening processes. We show that sequential screening pipelines exhibit new and surprising behavior where individuals can exploit the sequential ordering of the tests to ``zig-zag" between classifiers without having to simultaneously satisfy all of them. We demonstrate an individual can obtain a positive outcome using a limited manipulation budget even when far from the intersection of the positive regions of every classifier. Finally, we consider a learner whose goal is to design a sequential screening process that is robust to such manipulations, and provide a construction for the learner that optimizes a natural objective.

2 - Financial Networks: Modeling, Inference, and Strategy

Akhil Jalan, UT Austin, Austin, TX, United States

Financial networks help firms manage risk but also enable financial shocks to spread. Despite their importance, existing models of financial networks have several limitations. Prior works often consider a static network with a simple structure (e.g., a ring) or a model that assumes conditional independence between edges. We propose a new model where the network emerges from interactions between heterogeneous utility-maximizing firms. Edges correspond to contract agreements between pairs of firms, with the contract size being the edge weight.

We show that, almost always, there is a unique "stable network." All edge weights in this stable network depend on all firms' beliefs. Furthermore, firms can find the stable network via iterative pairwise negotiations. When beliefs change, the stable network changes. We show that under realistic settings, a regulator cannot pin down the changed beliefs that caused the network changes. Also, each firm can use its view of the network to inform its beliefs. For instance, it can detect outlier firms whose beliefs deviate from their peers. But it cannot identify the deviant belief: increased risk-seeking is indistinguishable from increased expected profits.

Finally, we model how firms can exploit information about their counterparties to gain better contract terms during iterative pairwise negotiations. In our formulation, any subset of firms can be strategic. We give a polynomial-time algorithm to find the Nash equilibria, if they exist, and certify their nonexistence otherwise. Simulations on OECD trade data between 46 large economies demonstrate that network outcomes are sensitive to these strategic negotiations.

3 - Improved Mechanisms and Prophet Inequalities for Graphical Dependencies

Vasilis Livanos, University of Chile, Santiago, Chile

Over the past two decades, significant strides have been made in stochastic problems such as revenue-optimal auction design and prophet inequalities, traditionally modeled with n independent random variables to represent the values of n items. However, in many applications, this assumption of independence often diverges from reality. Given the strong impossibility results associated with arbitrary correlations, recent research has pivoted towards exploring these problems under models of mild dependency.
In this work, we study the optimal auction and prophet inequalities problems within the framework of the popular graphical model of Markov Random Fields (MRFs), a choice motivated by its ability to capture complex dependency structures. Specifically, for the problem of selling n items to a single buyer to maximize revenue, we show that the max of SRev and BRev is an $O(\Delta)$-approximation to the optimal revenue for subadditive buyers, where Δ is the maximum weighted degree of the underlying MRF. This is a generalization as well as an exponential improvement on the $\exp(O(\Delta))$-approximation results of Cai and Oikonomou (EC 2021) for additive and unit-demand buyers. We also obtain a similar exponential improvement for the prophet inequality problem, which is asymptotically optimal as we show a matching upper bound.

TE21

Summit - 341

Public Health Applications in Decision Analysis

Invited Session

Decision Analysis Society

Chair: Jacob Jameson, Harvard University, Cambridge, MA, United States

1 - Improving Hypertension Care Cascades and Financial Risk Protection in Low- and Middle-income Countries

Dorit Stein, Harvard University, Boston, MA, United States, Stéphane Verguet

Purpose. This research estimates the distributional financial risk protection impacts of improving hypertension management across socioeconomic groups in low- and middle-income countries.

Methods. We developed a microsimulation model using nationally-representative, individual-level survey data from 44 low- and middle-income countries. We simulated improvements in the proportion of people with hypertension on treatment up to 80%. We applied relative risk reductions to baseline CVD risk estimates for untreated individuals randomly initiated on anti-hypertensives. We estimated the proportion that experienced catastrophic health expenditure (CHE), defined as direct (inpatient out-of-pocket costs) and indirect (income loss) costs exceeding a percentage of household consumption.

Results. Without improvements in hypertension management and assuming individuals seek inpatient care after a CVD event, 7.5% (11.9%) of people living with hypertension in the bottom quintile across countries would experience CHE at a 40% (25%) threshold. With improved treatment coverage, 6.2% (10.0%) of individuals in the bottom quintile would experience CHE at a 40% (25%) threshold. Including indirect costs averted slightly greater numbers of CHE cases across all quintiles, but especially among the top quintiles that would experience larger income losses due to hospitalization. Considering socioeconomic-based disparities in care-seeking, 1.5% (40% threshold) and 2.5% (25% threshold) of individuals in the bottom quintile would experience a CHE event.

Conclusion. Equitable coverage in hypertension treatment could avert CHE cases by averting CVD cases and hospitalizations. Unequal access to inpatient care may undervalue the potential financial risk protection benefits of improved CVD prevention among the poorest who are often less likely to seek care.

2 - "Time travel" calibration: an approach to multi-cohort disease model calibration with partially observed target data

Valeria Gracia Olvera, Stanford University, Stanford, CA, United States, Fernando Alarid-Escudero, Jeremy Goldhaber-Fiebert

Implementation of health policies at different points in time can impact populations consisting of multiple birth cohorts. The effects of these policies must be estimated by accounting for differences between prior exposures and future risks for each birth cohort. A common challenge is that empirical data often does not contain each birth cohort's complete lifetime exposures, risks, and outcomes.

We propose a simulation model-based approach to estimating past cohort trends and then extrapolate future trends of incidence and mortality using only the observed portions of cohorts' lifetime data. We illustrate this approach by examining diseases related to outdoor air pollution in Mexico City. We specified a continuous-time, age-period-cohort, multi-disease system of ordinary differential equations fitted to data on age- and sex-specific all-cause and cause-specific mortality, and disease prevalence from the Global Burden of Disease Study 2019 (1990-2019) with historical information on all-cause mortality from the Mexican government (1970-1990). A sequential calibration approach was used to estimate cohort trends in the model. Calibration began with cohorts observed from birth, recovering any time trends in disease incidence and mortality for early ages. These trends inform priors on early-age incidence and mortality for subsequent calibration of earlier birth cohorts, only observed in middle ages. This process is repeated for cohorts whose first observations are at progressively older ages. Once we "time travel" as far back as there is data available, we use the calibrated trends in middle and older ages to extrapolate future trends for later ages of more recent birth cohorts.

3 - Decision science in low-resource, data-deserts: A case study in global digital health

Sarah Skye Yoden, Harvard University, Cambridge, MA, United States

More than 1,000 digital health innovations have been funded by philanthropic agencies and piloted in low- and middle-income countries with the hopes of achieving improvements in health, and financial sustainability, which is typically defined as being financially supported by domestic tax revenues and/or consumers rather than donor funds. More than 2,000 pages of guidance, and a suite of different tools, exist to guide decision-makers on identifying which digital health interventions may be cost-effective. Yet despite these available resources and the widespread piloting of digital tools, normative agencies such as the WHO have noted a paucity of rigorous published research assessing the health impacts and cost-effectiveness of digital health interventions. This talk will provide an overview of digital health in low-resource settings today, a landscape of cost-effectiveness tools developed by the World Bank and others to assess digital health tools in low-resource settings, and perspectives from governments, implementers and funding agencies outlining why these cost-effectiveness tools are rarely, if ever, used. The talk will close with reflections on 2-3 small changes, and one big change, decision scientists could make to their digital health cost-effectiveness tools to increase uptake in the global digital health community, and potentially other communities that struggle with data deserts and few financial resources.

4 - Analyzing the impact of telehealth modalities on hospital readmissions

Emad Abed, Binghamton University , Binghamton , NY, United States, Alireza Boloori, Kathy Li, Rupinder Jindal

The COVID-19 pandemic accelerated the adoption of telehealth. While telehealth solutions were used in the early period post pandemic to ensure continuity of care, many believe they are here to stay in the near future. To this end, utilizing U.S. patients' claims data, we investigate the impact of various telehealth modalities (e.g., audio-video vs. audio only) on hospital readmission rates. We establish associations between modalities of telehealth interventions and 30-day hospital readmissions, offering insights for improving patient care, healthcare delivery, and optimizing resource utilization.

TE22

Summit - 342

Analytics in Action: Decision Excellence, Risk Modeling, and Pedagogy

Contributed Session

Chair: Michelle Song, University of Washington, Seattle, WA, United States

1 - Data-Driven Decision Excellence: Integrating Decision Analysis and Data Analytics in Business Cases for Informed Decision-Making

Alan Ferrandiz, Core Analitica LLC, Tampa, FL, United States

Originating from Harvard Business School in 1921, the concept of business cases has long provided a structured framework for evaluating strategic initiatives. As data volumes surge and data analytics technologies become ubiquitous, the conversion of raw data into actionable insights underscores the critical interaction between data experts and decision-makers.

This presentation revolves around the imperative for data professionals to effectively communicate with decision-makers throughout the process of transforming raw data into informed business decisions, specifically, on the pivotal role of grouping and aggregating raw data to create meaningful measures, and strategically select the most pertinent ones to serve as key performance indicators (KPIs), guided by their direct impact on business objectives.

Expanding on this basis, the presentation highlights the crucial role of articulating the narrative of effective business cases, starting with the delineation of precise business objectives derived from KPIs, to outline strategic initiatives aimed at achieving these objectives and a thorough numerical assessment of the associated costs and returns.

Additionally, the presentation also emphasizes the crucial role of dashboards in querying and visualizing KPIs, offering decision-makers real-time insights into organizational performance, as well as the importance of reports to effectively communicate the evolution of KPIs over time and the importance of storytelling with data in these reports.

By combining theoretical insights with practical examples, participants will develop a comprehensive understanding of the complexities associated with integrating decision analysis and data analytics into business cases. Furthermore, they will obtain practical strategies for articulating data-driven insights effectively to decision-makers.

2 - Exploring the Possible: Decision Support Using Heuristics in Stochastic Programming- Optimizing Force Structure Under Uncertainty

Aron Wing, United States Air Force Academy, Colorado Springs, CO, United States

A framework is introduced to heuristically explore a Stochastic Program's solution space between the optimist and the pessimist to provide better decision support to senior leaders. While previous work is in the field focuses on chance constraints or maintaining robust solutions, this work focuses on the overall probability that the unknown parameters, once realized in the future, will be detrimental to the solution. This work is a response to “The Case for Change, Optimizing the Air Force for Great Power Competition”, General Allvin, CSAF (Feb 2024) and “Accelerate Change or Lose”, Gen Brown, CSAF (Aug 2020). By applying this framework, Statistically Robust Non-Linear Programming/Optimization may provide meaningful analysis concerning the optimal combination of available platforms both now and for future airpower competition.

3 - A Hierarchical Approach for Risk Modeling of System Availability

Michelle Song, University of Washington, Seattle, WA, United States

Predicting part obsolescence is an important aspect of understanding system availability. Naval and defense industries have long-lived systems where accurate prediction of system availability is critical to fleet sustainment and system health and reliability. The data used in obsolescence prediction is focused on the part-level and the lack of subsystem-level data makes it difficult to leverage machine learning techniques to determine relationships between parts and systems. A new approach is needed to overcome the challenging characteristics of obsolescence or availability data. A hierarchical system availability logic tree (H-SALT) leverages the structure of reliability block diagrams (RBD) and fault trees (FT) to mathematical model system availability. A system availability tree is created from an existing RBD. Then the tree is analyzed using a top-down successive Boolean substitution method, which eliminates all intermediary subsystems between individual parts and the overall system. This method overcomes the issue of data scarcity and allows system-level availability to be determined using only part-level obsolescence data. A minimal cut set is used to determine the set of groups of parts and/or subsystems that will cause the overall system to become unavailable. H-SALT determines both system-level availability and the parts and/or subsystems that are critical to overall system availability.

TE23

Summit - 343

Supply Chain Learning and Sourcing

Contributed Session

Chair: Xin Xu, Shanghai Jiaotong University, Shanghai, N/A, China, People's Republic of

1 - The Formation Mechanism of Internal Bullwhip Effect: Evidence from Chinese Manufacturing Enterprises

Yi Yang, College of Economics and Management, China Agricultural University, Beijing, China, People's Republic of, Jianjun Lu, Jiawei Xu

The bullwhip effect not only occurs between upstream and downstream enterprises, but also exists within a single enterprise, which poses a great challenge to the orderly production and distribution. This study introduces an intra-firm perspective to investigate the impact of enterprise management styles and operational behaviors on internal bullwhip effect. Using a survey data from 304 Chinese manufacturing enterprises, the result showed that demand information distortion, order batch processing and rationing game can amplify the effects of IBE. Furthermore, although good management practices at the organization, process and information levels cannot directly diminish IBE, it can have an effect on IBE through the indirect effect of poor operational behavior. According to the fsQCA analysis, we obtained three configurations of basic preconditions that lead to high IBE. Finally, we discussed the implications of the research findings from both theoretical and practical dimensions.

2 - Financing and Sourcing Model Under Quality Improvement: Credit Guarantees and Equity Financing

Cheng-Feng Wu, National Taipei University, Taipei, Taiwan

This study discusses a financing and sourcing model under quality improvement in a two-echelon supply chain, utilizing Stackelberg game theory. It aims to identify the optimal sourcing strategy for brand manufacturers adopting either buyer credit guarantees or equity financing when considering supplier quality improvement. Brand manufacturers facilitate suppliers in acquiring operational funds for production and quality improvement, aiming to maximize the efficiency of supply chain operations. Improvements in production and quality by suppliers not only enhance their operations but also assist the end market, thereby strengthening the competitiveness of the supply chain. However, suppliers lacking funds face financial difficulties and operational risks, unable to secure bank loans independently. To resolve these financial issues, downstream brand manufacturers assist upstream suppliers by providing buyer credit guarantees or equity investments, ensuring the smooth operation of the supply chain. Findings from the results indicate that among suppliers, a self-selection mechanism allows those with superior operational performance to secure orders, while less efficient suppliers may be excluded from the supply chain. For manufacturers, the requirement to assume a higher assessment rate correlates with an increased default probability of suppliers.

3 - Joint Emission Reduction and Competition in Supply Chain

Xin Xu, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Qinghua Zhu, Wenbin Wang

The growing concern over climate change has led to heightened scrutiny of product carbon footprints, prompting collaborative efforts across supply chain entities to reduce emissions. This move aligns with regulatory measures aimed at reducing emissions across supply chains and is further bolstered by consumer demand for low-carbon products, which is driving firms to invest in greener practices. This study explores a supply chain with a supplier and two competing manufacturers under emission regulations, integrating consumer preferences for low-carbon products. It investigates both vertical collaboration in green investment between the upstream supplier and downstream manufacturers and horizontal competition among manufacturers targeting eco-conscious consumers under emission regulations. Our research delves into how market potential and regulatory constraints interact to shape the green investment along the supply chain. Until a certain market size threshold is reached, the supply chain prioritizes regulatory compliance. Beyond this threshold, there's a significant increase in joint investment, surpassing regulatory requirements. Furthermore, intensified competition in price consistently encourages green investment. However, the impact of competition on greenness varies according to market potential: it discourages green investment in contexts with low market potential but fosters more significant investment in scenarios with high market potential. We also find that information asymmetry regarding market potential reduces the supplier’s willingness to invest in green initiatives. Lastly, we emphasize the intricate environmental impact of emission regulations, highlighting how stricter regulations may lead to increased total carbon emissions, contingent upon market size.

TE24

Summit - 344

Logistics and Network Optimization Challenges

Contributed Session

Chair: Austin Saragih, Massachusetts Institute of Technology, 70 Amherst St, Cambridge, MA, 02142, United States

1 - Execution Time Estimation for Amazon Last Mile Route Planning

Bharath Veluri, Amazon, Hyderabad, India, Duo Zheng

Amazon's logistics operations are growing significantly year over year. One of the key challenges it faces is the complexity of the "last mile" of delivery—the final stage where packages are transported from the delivery station to customers' doorsteps. To address this, Amazon is focused on creating safe and sustainable route planning and execution processes. To assist in the route planning process, accurately estimating the length and duration of delivery routes is crucial, as inaccurate estimates result in sub-optimal routes leading to poor transporter experience. In this talk, we discuss a deep learning-based solution developed by Amazon for predicting accurate execution times at different granularities.

2 - Dynamic Resilience and Risk Propagation in Copper Supply Networks: A Complex Network Analysis

QIYAN LIU, University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Adriaens Peter

Copper, a critical mineral for renewable energy technologies, faces significant supply chain disruption risks due to its uneven global distribution. This research examines vulnerabilities within the copper industry's supply chain, which is highly interconnected from raw material extraction to finished goods production.

Using complex network theory, we construct a weighted, multi-tier supply chain network model at the firm level, based on real-world data from the FactSet Supply Chain Relationship database. We analyze the network's structural characteristics and employ degree-based removal scenarios to simulate risk propagation in the network, providing insights into the network's robustness and the potential for cascading failures.

Our study enriches the literature by offering a detailed firm-level analysis of the copper supply chain and employing dynamic risk transmission modeling to assess supply risks across different levels. We identify key vulnerabilities and propose strategies to enhance network resilience, such as diversfying suppliers and imporving communication channels. Through this comprehensive framework, we provide new insights into mitigating supply chain vulnerabilities in the copper industry, emphasizing the critical role of network structure and interdependencies in supply chain resilience. Our findings highlight the necessity for proactive risk management and strategic planning to ensure the stability and sustainability of the cooper supply chain in the face of global disruptions.

3 - UPS Sleeper Network: Routing and Scheduling

Ahmed El-Nashar, UPS, Alpharetta, GA, United States

UPS middle mile network plays a pivotal role in the transportation of parcels across a sequence of hubs, orchestrated by a fleet of trucks strategically stationed at depots dispersed throughout the network's expanse.

This logistical framework is designed to efficiently consolidate volume into full truckloads, with these trucks traversing between the network's locations and hubs, facilitating the seamless conveyance of truckloads from their origination points to their designated endpoints.

To optimize resource allocation and meet stringent delivery schedules, the trucks are operated by a two-driver team, commonly referred to as a sleeper team.

This operational framework encapsulates the vehicle routing and scheduling problem, wherein the primary aim is to devise a set of feasible vehicle schedules that cover the network's loads while minimizing operational costs and accounting for a set of federal regulations and operational constraints.

In this paper, a mathematical model is formulated to address an extended variant of the vehicle routing and scheduling problem, encompassing multiple depots while also incorporating considerations for teams assignments and load time windows.

Given the NP-hard nature of the problem under consideration and the inherent difficulty in solving large-scale instances with exact algorithms, a hybrid approach integrating meta-heuristics and exact methods is proposed. This approach seeks to leverage the runtime enhancements afforded by meta-heuristics while simultaneously harnessing the solution quality improvements provided by exact methods.

4 - Analyzing School Bus Electrification in Richmond, Virginia

Yi He, National Renewable Energy Laboratory, Golden, CO, United States

School buses are an essential component of the transportation infrastructure, serving as a lifeline for students across the globe. However, the widespread use of diesel school buses has raised concerns about the health impact on millions of students exposed to harmful emissions daily. Recognizing this issue, school districts worldwide are urgently seeking cleaner energy alternatives. Electric school buses emerge as an environmentally friendly and sustainable option, fostering a healthier environment for both students and communities. However, school bus electrification faces the challenges of high upfront cost, cumbersome charging management, and constraints from power grids. To help school bus operators address those challenges, this study presents a data-driven analysis for school bus electrification. This study considers a real-world school bus system in Richmond, Virginia, and develops a mathematical programming model to analyze the system design, charging strategies, and charging load profiles for the electrification scenario. The study evaluates different charging strategies based on model outcomes, aiming to optimize efficiency and effectiveness. Ultimately, this research generates electric school bus charging demand profiles under various scenarios, shedding light on the feasibility and implications of transitioning to electric-powered school buses.

5 - The Capacitated r-Hub Interdiction Problem with Congestion: Models and Solution Approaches

Sneha Bhatt, Management Development Institute, Gurugram, India, Sachin Jayaswal, Ankur Sinha

We study the r-hub interdiction problem under the case of possible congestion. Hub interdiction problems are modelled as attacker-defender problems to identify a set of r critical hubs from a set of p hubs, which when attacked, cause the maximum damage to the network restoration activities of the defender. In this work we consider that in addition to the routing cost, the defender also aims to minimize the congestion cost. Incorporating the congestion cost in the problem introduces non-linearity in the objective function of the interdiction problem, which makes the problem challenging to solve. To address this, we propose two alternate exact solution approaches. The first approach is an inner-approximation-based approach (IBA), which overestimates the convex non-linear objective function and provides an upper bound. A lower bound is obtained from solving the lower-level problem exactly corresponding to the upper bound solution. The upper bound is tightened using improved approximation with new points generated in successive iterations. In the second approach (referred to as SBA), the problem is reformulated as a second-order
conic program, which can be solved using an off-the-shelf solver. From our computational experiments on benchmark datasets (CAB and AP), we demonstrate the efficacy of both the proposed methods. However, IBA consistently outperforms SBA by a significant margin

6 - Modern Supply Chain Design Under Uncertainty

Austin Saragih, Massachusetts Institute of Technology, Cambridge, MA, United States

In this work, we present new three results in to design a supply chain network under uncertainty and modern lens. We first show which uncertainties we should resolve in the network. We then leverage contextual optimization from data to make data-driven facility decisions. We finally integrate policy and supply chain optimization and leverage approximation to solve this efficiently.

TE25

Summit - 345

JMP/SAS

Invited Session

Technology Showcase

1 - Building Classification Prediction Models for Image Recognition using Torch Deep Learning

Kevin Potcner, JMP Statistical Discovery, San Francisco, CA, United States

Images are rich with information. Contained in the thousands of pixels within an image are patterns in shapes, colors, and textures. These features often correspond to attributes that classify images into belonging to a particular category. Advancements in deep neural network modeling have been made providing analysts with a mechanism train models on a set of images and then apply that model to new images to predict the categories those new images belong to.

In this presentation, a statistical scientist from JMP software will demonstrate a new add-in tool -Torch Deep Learning- in JMP to model and classify images.

2 - Building and Solving Optimization Models with SAS

Rob Pratt, SAS, Cary, NC, United States

SAS offers extensive analytic capabilities, including machine learning, deep learning, natural language processing, statistical analysis, optimization, and simulation. SAS analytic functionality is also available through the open, cloud-enabled design of SAS® Viya®. You can program in SAS or in other languages – Python, Lua, Java, and R. SAS Analytics is also equipped with AI-enabled automations and modern low-code or no-code user interfaces that democratize data science usage in your organization and offer unparalleled speed to value.

OPTMODEL from SAS provides a powerful and intuitive algebraic optimization modeling language and unified support for building and solving LP, MILP, QP, conic, NLP, constraint programming, network-oriented, and black-box models. This showcase will include an overview of the optimization capabilities and demonstrate recently added features.

TE26

Summit - 346

Market Dynamics and Predictive Modeling

Contributed Session

Chair: Max Hill, Intel

1 - Stock Investment Strategies: Introducing SIRIF and SIRI for Enhanced Stock Return Monitoring and Trading Decisions

Dr. Jayanta Pokharel, Siena College, Latham, NY, United States

Monitoring returns through stock price prediction is a daunting task in the real world, given the multitude of non-numerical factors and the inherent noise in the market. Our current research endeavors to tackle this challenge by introducing two innovative tools: the stochastic-index-return-intensity-function (SIRIF) and the stochastic-index-return-indicator (SIRI). Utilizing the Power Law process, these tools are crafted to effectively monitor weekly index returns and assist investors in formulating trading strategies for buy-sell-hold decisions. By harnessing data from the Information Technology Sector Index of the S&P 500, SIRI offers valuable insights for generating trading signals, thereby streamlining investment planning and decision-making processes for investors.

2 - Beyond Accuracy: Navigating the Association between Hierarchical Forecasting and Decision Making

Mahdi Abolghasemi, University of Queensland, Brisbane, Australia

Time series forecasting often involves forecasting a collection of series that are organised in a group or hierarchical form, e.g., across geographical regions, brands, or categories. Hierarchical forecasting methods can potentially improve forecast accuracy and promise to generate coherent forecasts that are meant to be used for consistent decision-making across different levels of a hierarchy. However, it is not evident if consistent forecasts lead to consistent decisions, and if coherency is necessary for better decision-making in practice. Studies show that forecasts do not translate directly to decisions. We show how reconciled forecasts can lead to various decisions in hierarchical time series. We discuss how relaxing strict coherency of forecasts across the levels can impact the forecast accuracy and decision quality accordingly. We then propose a novel approach to generate harmonic forecasts by considering decisions across the hierarchy.

3 - Sufficient dimension reduction in vector autoregressive models for multiple time series

Jiaying Weng, Bentley University, Waltham, MA, United States, Yaser Samadi

Vector autoregressive (VAR) models are fundamental tools for analyzing multivariate time series data across diverse domains such as economics, finance, and climate studies. However, modeling high-dimensional time series data poses challenges due to the curse of dimensionality, especially when incorporating multiple time series and escalating model complexity. Sufficient dimension reduction (SDR) is a concept in statistics and machine learning aimed at finding a lower-dimensional subspace of the original feature space that preserves the relevant information for the target variable(s) or response variable(s). This paper explores the SDR in nonlinear vector autoregressive (NVAR) models, where the current state depends on multiple indices defined in past lags with an unknown relationship such that the variations in the data that are relevant to predicting the target variable(s) are retained, while irrelevant variability is removed or minimized. We propose a novel time series martingale difference divergence matrix (TS-MDDM) approach tailored for non-sparse estimation, albeit suitable only for low-dimensional scenarios. However, in the context of high-dimensional complexities where the dimensions grow rapidly by increasing the size of series and lag order, we design a sparse estimation procedure within our proposed MDDM method, leveraging a regularized optimization framework equipped with the LASSO penalty. We rigorously present theoretical foundations for both non-sparse and sparse estimators. Through simulations and real data analyses, we demonstrate the efficacy of our methodology in handling high-dimensional time series data within NVAR frameworks.

4 - Joint Prediction and Optimization: a Case Study of University Course Timetabling

Guisen Xue, Kent State University, KENT, OH, United States

Enhanced accuracy in demand forecasts enables practitioners to effectively strategize and optimize subsequent operations, aiming to maximize profits or minimize costs. Therefore, demand forecasting holds a pivotal position in management practices. Typically, the optimization method is used to generate decisions, while a machine learning model can predict the key parameters, such as demands, of the optimization model in advance. However, this predict-then-optimize paradigm becomes inappropriate when the predicted parameters are affected by the decision variables in optimization. In such instances, the joint prediction and optimization method is preferred for smarter decisions. This study constructs a mixed integer non-linear programming model for this framework, applicable to various real-world optimization scenarios. The focal application context is university course timetabling problem (UCTP), where factors such as assigned instructors, timeslots, and delivery methods impact estimated student course demands, determined during the optimization phase. The predict-then-optimize paradigm and the joint prediction and optimization are analyzed and compared in the UCTP in a department with multiple disciplines at a comprehensive public university. The results demonstrate the superiority of the objective over the predict-then-optimize paradigm and that of the solution complexity over other joint prediction and optimization methodologies.

5 - Modeling markets as dynamical systems

Max Hill, Intel, Ridgefield, WA, United States, Aparna Komarla

Autoregressive forecasting techniques are useful for short-term predictions but offer little guidance in unfamiliar or rapidly changing markets. Despite the growing complexity of modern market dynamics, customer purchasing behavior can be understood and anticipated if analyzed through a dynamical systems lens similar to those used to model diseases, populations, and many other physical systems. In the context of market modeling, it is helpful to think about interrelated “nodes" that could be customer or product segments. Once the nodes are defined, one can write differential equations for the size of the nodes over time as functions of customer population sizes and purchasing behavior drivers (e.g. device age, quality, price, etc. of the various products). In many cases, good estimates of the equation parameters can be obtained through global constrained optimization. While some drivers of demand are less predictable (including “black-swan” events and major technological disruptions), we will show how the likely consequences of these events can be determined more quickly than might otherwise be possible, resulting in a more rapid response or greater organizational readiness for future market changes. An implementation of this methodology is being developed at Intel and is being applied to the PC market. In this talk, we will discuss this approach, our initial results, and how the approach can be extended to other industries.

TE27

Summit - 347

Emerging Strategies in Platform Competition and Innovation Management

Invited Session

New Product Development

Chair: Junghee Lee, University of Notre Dame, Notre Dame, IN, United States

1 - Cooperative OR Non-Cooperative New Product Development

Sara Rezaee Vessal, ESSEC Business Scchool, Paris, France, Timofey Shalpegin

In a product development project, the buyer might have some uncertainties about the characteristics or requirements of the product at the time of awarding the contract to a supplier. To include the possibility of revising and/or extending the scope of the contract, the buyer can consider coordination or renegotiation with suppliers. Our goal is to track the effect of cooperation between the buyer and the supplier on the outcome of the project. We consider three settings, with no-, partial, and full cooperation, considering sequential game with risk-neutral players and analyse their equilibrium strategies. We find that buyer's barraging power or share of cost and the development cost impact players' optimal strategy. We also study how their strategy is impacted by projects' and environment' characteristics.

2 - Access to Science in the Town Square: The Role of Social Media in Science Communication

Yang Yang, University of Notre Dame, Notre Dame, IN, United States, Tanya Tian, Brian Uzzi, Benjamin Jones

Science is increasingly communicated to the public through social media, raising significant questions about the role that social media serves in how scientific knowledge is consumed in public domains. Here, we engage in systematic investigation combining large-scale datasets and establish several core findings. First, social media massively expands the communications of science beyond mainstream media outlets. Second, despite prevailing skepticism about social media in the age of misinformation, we find that social media covers high-quality works in the domain of science. Third, in contrast to mainstream media that emphasizes eminence in the coverage of science, social media enables more diverse voices to be heard more evenly across dimensions of field, institutional and journal prestige, and gender. Lastly, at the scientist level, there is remarkable alignment between impact in social media and impact within science. Contrary to prevalent concerns about social media across domains, these findings highlight that social media presents diverse scientific communication, communicating a greater volume and range of high-quality scientific works while drawing more evenly from a broader set of voices on distinctive dimensions.

3 - Token Incentives and Platform Competition: A Tale of Two Swaps

Xiaofeng Liu, Baruch College - The City University of New York, New York, NY, United States, Wei Chen, Kevin Zhu

In the digital economy, platforms compete intensively to attract users, often using expensive subsidies. Token incentives emerge as a new incentive for crypto platforms to attract participants. In this paper, we study the impacts of cost-effective token incentives in the competition between two decentralized exchange platforms. The entrant platform, Sushiswap, launched token incentives to lure liquidity providers away from the incumbent, Uniswap, who then retaliated with its own token incentives. Our analysis shows that the incumbent's own token incentives attract more liquidity to the platform. Yet, surprisingly, the token incentives from its competitor also bring more liquidity and trade volume to the incumbent. To understand the mechanisms behind the impact, we examine the role of network effects and information diffusion. The findings of this study provide theoretical insights and practical guidelines for platform managers on the design of token incentives in platform competition.

TE28

Summit - 348

Service Systems Operations

Contributed Session

Chair: Tianyu Pan, University of Florida, Gainesville, FL, United States

1 - Retail Investment Under Aggregate Fluctuations

Zhide Wang, Southern Methodist University, Dallas, TX, United States, Yanling Chang, Nathan Yang, Alfredo Garcia

Although local market conditions are normally considered to be the key drivers of forward-looking and strategic retail chain expansion (or contraction), broader aggregate states and their movements (i.e., aggregate fluctuations) might also be of relevance; however, the extent to which these aggregate states actually impact retail investment decisions remains unknown. This paper develops a new framework for modeling, identifying, and estimating dynamic investment games that accommodates for aggregate states. In particular, firms in our model make strategic investment decisions based on beliefs they have about the evolving aggregate state. With this novel model framework, we present a set of identification results, as well as a likelihood-based estimation methodology. We then conduct an empirical case study about fast food industry dynamics. Analysis of estimated models reveals that while aggregate states do indeed impact retail investment, these effects are heterogeneous, as some of the retail chains are noticeably more/less sensitive to the aggregate state. Furthermore, the estimated model demonstrates a connection between heightened volatility in the aggregate fluctuations and dampened retail investment. Finally, the relationship between volatility of the aggregate state and investment has the potential to further reinforce the market presence of already dominant firms.

2 - Integrating Reinforcement Learning Algorithms into Markov Decision Process for Hotel Dynamic Pricing Optimization

Aysajan Eziz, Ivey Business School, Western University, London, ON, Canada

This research article presents a novel approach for dynamic pricing in the hotel industry by integrating reinforcement learning (RL) algorithms into a Markov decision process (MDP) formulation. The model optimizes booking prices for multi-day stays, taking into account room availability and demand fluctuations. The RL algorithm helps find optimal pricing strategies, yielding higher revenues and minimizing losses during high-demand periods, as validated by computational experiments. While primarily focused on the hotel industry, the findings can be applied to other sectors with capacity constraints and variable demand, like transportation and e-commerce, providing valuable insights for dynamic pricing and revenue management.

3 - Strengthening Financial Sustainability of Community Mini-Grids: A Capacity Utilization Optimization Approach

Madhav Sharma, Indian Institute of Technology Kanpur, KANPUR, India

While community minigrids (MGs) in low- and middle-income countries (LMICs) hold promise for expanding rural electrification, ensuring their long-term financial viability remains a critical hurdle. This paper proposes a strategic framework to enhance the financial sustainability of these MGs by optimizing their capacity utilization factor (CUF). We introduce a unique approach that integrates energy-intensive processes, specifically crop processing activities (e.g., milling, drying), with periods of low CUF. This innovative approach not only increases CUF but also reduces operational costs, ultimately leading to lower tariffs for end-users. Through a comprehensive case study and rigorous analysis, we demonstrate the effectiveness of this strategy in meeting energy demands while ensuring the long-term viability of these MG systems. Our research emphasizes the crucial role of community mini-grids in fostering sustainable rural development and provides valuable insights for policymakers, practitioners, and stakeholders in the renewable energy and agricultural sectors.

4 - Information Leakage Prior to SEC Form Filings—Evidence from TAQ Millisecond Data

Weiting Hong, Yale School of Management, New Haven, CT, United States, Steven Ho, Mingrui Zhang

We investigate the stock price movements prior to the publication of publicly-listed firms' SEC form filings. By analyzing the time-stamps of all SEC form filings as well as the stock prices in the 30-minute interval pre- and post-publication, utilizing the TAQ millisecond data, we find strong evidence indicating that price movements in the 30-minute pre-announcement intervals can strongly predict the direction of the price movement for the post-30-minute interval, in that if the stocks are ranked into 5 portfolios based on the price run-up prior to filing release, the events with the highest run-up would also have the highest price increase post filing release, and vice versa (excluding form-types with directional implications, e.g., Forms 13G, will strengthen the results, though their inclusion does not invalidate the statistical significance of the overall results). Regressing pre-announcement return on buyer-initiated (classified according to Lee and Ready, 1991) minus seller-initiated trade volume fraction yields a positive and highly statistically-significant coefficient. The results are not explained away by momentum, common controls, the removal of filings potentially complicated by recent firm-specific-news, or the presence of 5-minute-interval fixed effects, and they remain after the SEC's server fix in March 2015 and after the prosecution of hackers trading on information obtained from filing agents in September 2020. To the best of our knowledge, this is the first paper that systematically documents the pervasiveness of the phenomenon by simply investigating prices in a model-free setup, though our results are also corroborated through other angles beyond prices (e.g., buyer-initiated proportions).

TE29

Summit - 420

Computational Integer Optimization and Applications

Invited Session

OPT: Integer and Discrete Optimization

Chair: Ozge Aygul, Worcester Polytechnic Institute, WORCESTER, MA, 01609, United States

1 - Scheduling Cybersecurity Mitigations to Delay Attacker Projects

Ashley Peper, University of Wisconsin-Madison, Madison, WI, United States, Jim Luedtke, Laura Albert

We study the problem of deploying cybersecurity mitigations to delay the completion times of a set of attacker projects. Existing literature studies how to select a portfolio of mitigations subject to a budget constraint to optimally delay such attack projects, but lacks consideration of the time it takes to implement the mitigations. We assume attackers are working on their projects at the same time the defender is working to complete mitigations to delay those attacks. Our proposed model captures the importance of scheduling mitigations, since to be useful they must be completed before an attacker exploits a vulnerability. This paper introduces an integer programming model of this problem that

combines a defender's mitigation scheduling problem with the attackers' project completion problem. We consider alternative solution methods and examine the benefit of the proposed model.

2 - New Explorations in Integer Optimization for Minimal Egyptian Fraction Representations

Sepideh Sedghi, Worcester Polytechnic Institute, WORCESTER, MA, United States, Andrew Trapp

Egyptian fractions are a mathematical challenge of antiquity concerned with representing fractions as sums of reciprocals. One fundamental open question is: What is the fewest number of terms required to represent a given fraction? We propose an innovative integer optimization model to address this question. Our solution approach integrates a custom-designed Branch and Bound algorithm coupled with an exact solver for the LP relaxations. This approach ensures result accuracy through exact computation and overcomes roundoff errors encountered with traditional solvers. We incorporate a number of computational enhancements including valid inequalities to accelerate the solution process. We evaluate our model with thousands of instances and show our approach on average can quickly identify minimal representations, while demonstrating improved accuracy over traditional solvers and performance gains over alternative methods.

3 - Strategies for Dual Overhead Crane Scheduling

Deniz Nigar Yoltay, Ozyegin University, Istanbul, Turkey, Erhun Kundakcioglu

In this study, we examine scheduling algorithms for two overhead cranes operating on the same parallel tracks. Efficient scheduling of these dual cranes is crucial for improving productivity and functionality in manufacturing and logistics facilities. The problem involves tasks that need to be completed at the beginning of a planning horizon, with each task associated with a specific origin-destination pair and required time. As the cranes cannot pass over each other horizontally, conflicting assignments can lead to potential crane stops. The objective is to maximize the sum of priorities for scheduled tasks within the given timeframe.

We introduce a Dynamic Programming (DP) approach that systematically evaluates all feasible schedules. However, the computational complexity of the DP algorithm restricts its use in large-scale scheduling tasks. To overcome this limitation, we propose two modifications: Restricted Dynamic Programming (RDP) and Approximate Dynamic Programming (ADP). RDP reduces the state space at each stage, while ADP approximates upcoming crane conflicts, thereby decreasing solution time.

Finally, we propose a Time-Space Network approach, which represents crane actions across spatial locations and discrete time intervals. In this model, nodes denote location-time pairs, and arcs represent crane movements and task processing. We evaluate the performance of our algorithms across various instance sets and parameters, and our numerical studies demonstrate the effectiveness of the Time-Space Network approach.

4 - Dynamic Course Scheduling for Strategic University Scaling

Özge Aygül, Worcester Polytechnic Institute, WORCESTER, MA, United States, Shima Azizi, Andrew Trapp

Universities face significant challenges in long-term planning due to fluctuating student enrollments, changing course demands, and the need to optimize resource utilization. Traditional methods often fall short in effectively managing these complexities, leading to inefficiencies and unmet educational goals. In this study, we present a novel mathematical optimization formulation designed to assist universities in making informed decisions for long-term planning. Our approach addresses the critical need to manage course section splits and merges, enabling more effective planning over extended periods. First, we estimate major distributions and course demand for the next four years, which serve as key parameters for our model. The mixed integer optimization formulation then assigns course sections to time patterns and locations, splitting and merging sections, and allows capacity extension as necessary. This model incorporates multiple competing objectives, including minimizing the splitting of course sections, honoring faculty preferences, and reducing unused classroom space. By simulating various growth scenarios, we evaluate the impact of these factors on university scheduling. Our formulation represents the first optimization model specifically tailored for long-term student body planning in higher education. The results demonstrate how strategic adjustments can enhance scheduling efficiency and resource utilization, offering valuable insights for university administrators and policymakers. This innovative approach offers a scalable solution for universities aiming to improve planning and adapt to changing student needs.

TE30

Summit - 421

Randomized Methods for Combinatorial Scientific Computing

Invited Session

OPT: Computational Optimization and Software

Chair: Sven Leyffer, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: Dominic Yang, Argonne National Laboratory, Lemont, IL, United States

1 - Randomized Construction of Approximate Optimality Certificates for Combinatorial Problems

Dominic Yang, Argonne National Laboratory, Lemont, IL, United States

Primal algorithms for combinatorial optimization problems generally proceed by repeatedly augmenting a feasible point using a collection of moves. When this set of moves is suitably expansive enough that if we cannot find an augmenting move, we are at the optimum point, we have an optimality certificate. For select problems, such collections are well-known, for example augmenting paths in the maximum flow problem or negative cost cycles in the minimum cost flow problem. For general integer programs, there have existed several notions of optimality certificates, e.g., Groebner bases and Graver bases, but these sets are generally too large to be of practical use. In this talk, we will consider the problem of relaxing the optimality requirement of an optimality certificate and attempt to establish a weaker criterion so that the produced sets are more manageable. We will present a reinforcement learning algorithm for learning a partial construction of these larger optimality certificates adapting the classic algorithms for producing Groebner bases, Buchberger's algorithm. This method will be based on randomized variants of Buchberger's algorithm to effectively learn how to construct a manageable set of moves.

2 - Sampling-based Progressive Hedging for Large Scale Stochastic Programming

Di Zhang, University of Southern California, Los Angeles, CA, United States, Suvrajeet Sen

The Progressive Hedging Algorithm (PHA) is a cornerstone for solving large-scale stochastic programming (SP) challenges. However, its traditional implementation is hindered by several limitations, including the requirement to solve all scenario subproblems in each iteration, reliance on an explicit probability distribution, and a convergence process that is highly sensitive to the choice of the so-called “prox” parameter. This paper introduces a sampling-based PHA that aims to overcome these limitations. Our approach employs a dynamic sampling process for the number of scenario subproblems solved per iteration. The new algorithm incorporates a) an adaptive sequential sampling process for determining sample sizes, b) a stochastic conjugate subgradient method for direction finding, and c) a line-search technique to update the dual iterates. Experimental results demonstrate that this novel algorithm not only addresses the bottlenecks of the conventional PHA but also potentially surpasses its scalability, representing a substantial improvement in the field of stochastic programming.

3 - Randomization for Fiber Sampling in Discrete Optimization

Ivan Gvozdanovic, Illinois Institute of Technology, Chicago, IL, United States, Sonja Petrovic, Miles Bakenhus

We consider the problem of optimizing some (linear) objective function over lattice points in a polytope. For large, sparse, high-dimensional problems, one natural way to tackle the problem is using randomization, for example Markov Chain Monte Carlo (MCMC) algorithms, to search the feasible region. A priori, any MCMC algorithm requires a Gr\"obner or Graver basis as input in order to guarantee connectivity of the chain. These bases act as optimality certificates and are central to non-linear algebra, but are generally too difficult to compute. In contrast, lattice bases are much easier to compute using linear algebra, and their linear combinations can be used to construct the larger bases and discover new feasible points in the polytope. The main open question is how to find these combinations in order to improve upon an existing feasible solution with respect to some objective function. This talk will review recent work on using reinforcement learning and biased randomized search algorithms to explore the lattice points in a polytope. We will discuss the application of the method to discrete optimization problems. This is joint work with Miles Bakenhus and Sonja Petrovi\'c.

4 - Construction of Hierarchically Semi-Separable Matrix Representation using Adaptive Johnson-Lindenstrauss Sketching

Yotam Yaniv, Lawrence Berkeley National Lab, Berkeley, CA, United States

In this talk we present an adaptive, partially matrix-free, Hierarchically Semi-Separable matrix construction algorithm that leverages randomized sampling via matrix sketching. We focus on adaptively applying the Johnson-Lindenstrauss transform (JLT) to approximate the range of our compressible operator. This algorithm is an extension of Gorman et al. [SIAM J. Sci. Comput. 41(5), 2019] which leveraged Gaussian sketching operators. We first present theoretical work which justifies this extension via concentration bounds. Then, we discuss the implementation details of applying Sparse JLT efficiently. Finally, we demonstrate experimentally that using structured or sparse random matrices instead of Gaussians leads to a speedup of the HSS construction implementation in the STRUMPACK C++ library while maintaining similar relative error in both the parallel shared memory and parallel distributed memory settings.

TE31

Summit - 422

Advances in Data-Driven Distributionally Robust Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Tianyu Wang, Columbia University, New York, NY, United States

Co-Chair: Henry Lam, Columbia University, New York, NY, United States

1 - Stability Evaluation via Distributional Perturbation Analysis

Jiashuo Liu, Tsinghua University, Beijing, China, People's Republic of, Jose Blanchet, Jiajin Li

The performance of learning models often deteriorates when deployed in out-of-sample environments. To ensure reliable deployment, we propose a stability evaluation criterion based on distributional perturbations. Conceptually, our stability evaluation criterion is defined as the minimal perturbation required on our observed dataset to induce a prescribed deterioration in risk evaluation. In this paper, we utilize the optimal transport (OT) discrepancy with moment constraints on the (sample, density) space to quantify this perturbation. Therefore, our stability evaluation criterion can address both data corruptions and sub-population shifts—the two most common types of distribution shifts in real-world scenarios. To further realize practical benefits, we present a series of tractable convex formulations and computational methods tailored to different classes of loss functions. The key technical tool to achieve this is the strong duality theorem provided in this paper. Empirically, we validate the practical utility of our stability evaluation criterion across a host of real-world applications. These empirical studies showcase the criterion’s ability not only to compare the stability of different learning models and features but also to provide valuable guidelines and strategies to further improve models.

2 - Hedging Complexity in Generalization via a Parametric Distributionally Robust Optimization Framework

Tianyu Wang, Columbia University, New York, NY, United States, Garud Iyengar, Henry Lam

Empirical risk minimization (ERM) and distributionally robust optimization (DRO) are popular approaches for solving stochastic optimization problems that appear in operations management and machine learning. Existing generalization error bounds for these methods depend on either the complexity of the cost function or dimension of the random perturbations. Consequently, the performance of these methods can be poor for high-dimensional problems with complex objective functions. We propose a simple approach in which the distribution of random perturbations is approximated using a parametric family of distributions. This mitigates both sources of complexity; however, it introduces a model misspecification error. We show that this new source of error can be controlled by suitable DRO formulations. Our proposed parametric DRO approach has significantly improved generalization bounds over existing ERM and DRO methods and parametric ERM for a wide variety of settings. Our method is particularly effective under distribution shifts and works broadly in contextual optimization. We also illustrate the superior performance of our approach on both synthetic and real-data portfolio optimization and regression tasks.

3 - Robust Decisions with Heavy Tailed Data: Statistics and Algorithms

Bart Van Parys, Massachusetts Institute of Technology, Cambridge, MA, United States, Bert Zwart

Most data-driven decisions formulations in the literature explicitly assume light-tailed or even bounded distributions. However, many real-world phenomena exhibit heavy-tailed distributions, characterized by rare but extreme events with may have significant impact. In this work we particular investigate a setting in which the data distribution merely has a bounded α-moment with $\alpha\in (1, 2)$. In particular this means that the variance of the data generating distribution may not any longer be bounded. In such heavy tailed setting classical approaches based on for instance the Wasserstein distance may fail due to their inherent sensitivity to outliers. Perhaps surprisingly, we show that despite that the variance may be infinite standard empirical variance formulations still retain some of their statistical efficiency properties which are well known in the $\alpha\in [2, \infty)$ regime. Finally, we indicate that a judiciously scaled Kullback-Leibler formulation is statistical efficient as well as computationally tractable.

4 - Master Surgery Scheduling Problem Considering Uncertainty in Length of Stay in Downstream Units

Nadia Lahrichi, Polytechnique Montreal, Montreal, QC, Canada, Erfaneh Nikzad

This study introduces an innovative mathematical model for the master surgery scheduling problem, a critical aspect of operating room planning at the tactical level. Failing to consider resource availability in downstream units, such as wards and ICU beds, which patients require during post-surgery recovery, can lead to impractical scheduling, resulting in surgery cancellations or premature discharge, ultimately impacting patient satisfaction. The length of stay post-surgery, both in the surgical unit and ICU, is uncertain and influences bed availability. Therefore, this study incorporates post-surgery length of stay as stochastic parameters, defined within a limited set of scenarios. A two-stage stochastic programming model is employed to address uncertainties. The results show that considering uncertainty affects surgery scheduling in the planning horizon and significantly reduces the number of canceled surgeries due to lack of capacity in wards and ICU beds. In this study, patients are categorized into different classes for each service based on their diagnoses. Additionally, to ensure that operating room time capacities are respected, patterns are generated, indicating all possible combinations of classes that can take place in the same room in one day. Furthermore, a two-phase matheuristic algorithm based on a fix-and-optimize procedure is proposed to efficiently solve the model. In the first phase of the developed algorithm, an initial solution is generated, and then in the second stage, the initial solution is improved using the fix-and-optimize approach. The results demonstrate that the proposed matheuristic can obtain high-quality solutions in reasonable time, especially when the number of scenarios is large.

TE32

Summit - 423

Recent Advancements in Accelerated Optimization

Invited Session

OPT: Global Optimization

Chair: Mikhail Bragin, University of California, Riverside, Marina del Rey, CA, United States

1 - Polyhedral Analysis of Quadratic Optimization Problems with Stieltjes Matrices and Indicators

Peijing Liu, University of Southern California, Los Angeles, CA, United States, Alper Atamturk, Andres Gomez, Simge Kucukyavuz

We consider convex quadratic optimization problems with indicators on the continuous variables. In particular, we assume that the Hessian of the quadratic term is a Stieltjes matrix, which naturally appears in sparse graphical inference problems and others. We describe an explicit convex formulation for the problem by studying the Stieltjes polyhedron arising as part of an extended formulation and exploiting the supermodularity of a set function defined on its extreme points. Our computational results confirm that the proposed convex relaxation provides an exact optimal solution and may be an effective alternative, especially for instances with large integrality gaps that are challenging with the standard approaches.

2 - A linearly convergent Gauss-Newton subgradient method for ill-conditioned problems

Tao Jiang, Cornell University, Ithaca, NY, United States, Damek Davis

We analyze a preconditioned subgradient method for optimizing composite functions $h\circ c$, where h is a locally Lipschitz function and c is a smooth nonlinear mapping. We prove that when c satisfies a constant rank property and h is semismooth and sharp on the image of c, the method converges linearly. In contrast to standard subgradient methods, its oracle complexity is invariant under reparameterizations of c.

3 - Indefinite Quadratic Programs and Complementarity Constraints by a Progressive MIP Method

Xinyao Zhang, University of Southern California, Los Angeles, CA, United States, Shaoning Han, Jong-Shi Pang

Indefinite quadratic programs (QPs) are known to be very difficult to solve to global optimality, so are linear programs with linear complementarity constraints. Treating the former as a subclass of the latter, we present a progressive mixed integer linear programming method for solving a general linear program with linear complementarity constraints (LPCC). Instead of solving the LPCC with a full set of integer variables expressing the complementarity conditions, the presented method solves a finite number of mixed integer subprograms by starting with a small fraction of integer variables and progressively increasing this fraction. After describing the PIP (for progressive integer programming) method and its various implementations, we demonstrate, via a set of computational experiments, the superior performance of the progressive approach over the direct solution of the full-integer formulation of the LPCCs. It is also shown that the solution obtained at the termination of the PIP method is a local minimizer of the LPCC, a property that cannot be claimed by any non-enumerative method for solving this nonconvex program. In all the experiments, the PIP method is initiated at a feasible solution of the LPCC obtained from a nonlinear programming solver; such a solution is always improved by PIP in the subsequent iterations. Thus, the PIP method can improve a stationary solution of an indefinite QP, something that is not likely to be achievable by a nonlinear programming method.

4 - A Machine Learning Approach for Rank-1 GMI Cuts

Berkay Becu, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey, Qiu Feng, Alinson Xavier

Many optimization problems involve solving similar MILP instances repeatedly within a short time window. These instances typically have the same set of variables and constraint matrix but differ in objective coefficients and right-hand-side (RHS) values. Rank-1 GMI cuts are an important class of cutting-planes, being numerically stable compared to higher rank GMI cuts, which are generated from a row of the simplex tableau corresponding to a fractional basic variable. Heuristics that provide effective approximation of the (rank-1) GMI closure are not efficient enough to be implemented within a solver. RHS changes imply problem changes which prevents learning and using cutting planes from a similar instance and ensure its validity. Since rank-1 GMI cuts can be characterized by the basis and the corresponding tableau row needed to generate it, we try to learn which fractional rows produce a good cut. Our conjecture is that if two instances are very similar, in terms of objective or RHS, then the rank-1 GMI closure can be obtained approximately with the same set of bases and fractional rows. This conjecture is the motivation for using a simple k-neighborhood learning algorithm to efficiently select Gomory mixed-integer (GMI) cuts from historical data of similar instances. We generated families of instances by perturbing the right-hand-side and objective functions of MIPLIB 2017 Benchmark. On instances where cuts generated by our implementation of GMIC closure approximation improves Gurobi's performance, cuts learnt from historical data were able to reduce running time of default Gurobi by 22% on average.

TE33

Summit - 424

Data-driven Stochastic Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Yuang Chen, Georgia Tech Shenzhen Institute, Tianjin University, Shenzhen, N/A

1 - Exploring Statistical Properties of Structured Optimal Policies Using Bootstrap Methods in Data-Driven Dynamic Programming

Yuang Chen, Georgia Tech Shenzhen Institute, Shenzhen, China, People's Republic of, Shiming Deng, Wanpeng Wang

Dynamic programming (DP) is fundamental in solving multi-period decision-making problems, where structurally optimal policies, such as the (S,s) policy in inventory control and the scheduling threshold policy in maintenance planning, are commonly used. However, analyzing the distributional properties of the estimated policy and corresponding total cost becomes challenging in the presence of unknown seasonal data. This challenge is primarily due to two factors: the recursive nature of the DP solution, which leads to error propagation, and the dependence of the estimated policy on all data samples, causing the total costs correlated. To address these challenges, we first transform the optimality conditions across multiple periods into a system of equations and apply the Z-Estimator Master Theorem to ascertain the asymptotic properties of policy parameters. Then, we introduce bootstrap methods to compute the distribution of the estimators for the policy parameters and their associated total costs, and prove the validity of these methods. Additionally, we also develop techniques for estimating the necessary sample size to achieve a specified precision in total cost estimates. Numerical studies in the contexts of inventory control and maintenance planning demonstrate that our methods outperform existing approaches in estimating statistical properties, thereby providing more reliable and accurate results.

2 - Data-Driven Contextual Inventory Policy with Censored Demand Based on Kaplan-Meier Estimators

Wanpeng Wang, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

We investigate the data-driven inventory problem faced by retailers of perishable products who lack knowledge of the true demand distribution, but have historical sales data (demand censored by available stock) along with a large number of features potentially related to demand. We propose a non-parametric method named censored kernel-weights optimization (CKO) and a parametric method named censored empirical risk minimization (CERM). Both CKO and CERM integrate optimization with demand estimation using appropriate contextual Kaplan--Meier estimators. We prove that the inventory policies computed using CKO and CERM converge in probability to the true optimal policy, considering demand censoring and observed features. Additionally, we characterize the finite-sample performance bounds on out-of-sample costs. To handle scenarios with a large number of features, we develop a high-dimensional CERM (H-CERM) algorithm considering adaptive Lasso penalty terms. We prove that, as the sample data size increases, H-CERM exhibits not only policy consistency, but also model selection consistency. That is, the probability of accurately identifying the true model with the right features approaches one. Numerical experiments confirm that our proposed algorithms outperform existing methods by converging to the true optimal policies with significantly lower out-of-sample total costs. Furthermore, our H-CERM algorithm has a higher probability of selecting the right features than existing approaches.

3 - Decisional SVM with Operational Constraints

Xianghui C, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Shiming Deng

We propose a decisional classifier framework that addresses operational constraints in personalized decision-making scenarios. Unlike traditional classifiers that solely focus on accurate classification by minimizing misclassification probability, our approach, Decisional SVM (DSVM), aims to maximize(minimize) the profit(cost) of personalized decisions based on observed features. Furthermore, DSVM incorporates operational constraints commonly associated with specific decision-making processes, which are often overlooked by traditional classifiers. To accommodate both constrained and unconstrained scenarios, we develop exact mixed-integer programming models for DSVM based on observed data. We also introduce approximate linear programming models tailored for constrained and unconstrained DSVM to present a balance between profitability and computational efficiency. For unconstrained DSVM, we demonstrate the universal consistency of both the exact and approximate models in relation to the theoretically optimal cost-oriented classifier. For operationally constrained DSVM, we provide a convergence rate for the feasible region defined by empirical constraints to the true feasible region. Through experiments conducted on synthetic and real-world datasets, we compare the performance of DSVM, approximate DSVM, and SVM, and analyze the circumstances where DSVM outperforms SVM.

4 - Best solution for Data-driven Robust Optimization with Chance Constraints

Shiming Deng, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

TBD

TE34

Summit - 425

Emerging Trends in Mathematical Modeling and Optimization

Contributed Session

Chair: Hongcheng Liu, University of Florida, Gainesville, FL, United States

1 - Variable selection in convex nonparametric least squares via structured Lasso penalty: An application to Swedish electricity market

Zhiqiang Liao, Aalto University, Espoo, Finland

We study the problem of variable selection in convex nonparametric least squares (CNLS). Whereas the Lasso is a popular technique for simultaneous estimation and variable selection, its performance is unknown in nonparametric convex regression problems. In this work, we investigate the performance of the Lasso regularized CNLS estimator in a high-dimensional setting and propose new Lasso methods by considering the unique structure of the subgradients. The proposed Lasso regularization approaches are further extended to doubly penalized methods when the regression model is assumed to be semi-nonparametric. The doubly penalized approach is applied to predict the total cost of Swedish electricity distribution system operators and empirically analyze the effects of variable selection. The results from the simulation and application confirm that our proposed methods perform favorably, while generally leading to sparser models, relative to the other penalties used in the CNLS estimator.

2 - A Novel Multivariate Grey Forecasting Model with Hausdorff Fractional Derivative and Its Applications

Gazi Duman, University of New Haven, West Haven, CT, United States, Elif Kongar

This research proposes a new forecasting method called the Hausdorff fractional NBGMC(p,n), which is an extension of the original nonlinear grey Bernoulli model with convolution integral, NBGMC(1,n). The new approach incorporates the Hausdorff fractional accumulation operator, providing more flexibility. Furthermore, recurrence relation of the binomial in the discrete solution provides simpler computation due to the elimination of the Gamma function calculation. The Jaya Algorithm is also employed to optimize the parameters of this new model, enhancing its ability to adapt. To validate the model's effectiveness, a case study on forecasting electronic waste is conducted. The results demonstrate that the proposed fractional model outperforms existing grey models in terms of its accuracy and predictive capability.

3 - Theoretical Approximation Ratios for QAOA on 3-Regular Max-Cut Instances at Depth p=1

Reuben Tate, Los Alamos National Laboratory, Los Alamos, NM, United States, Stephan Eidenbenz

We generalize Farhi et al.'s 0.6924-approximation result technique of the Max-Cut Quantum Approximate Optimization Algorithm (QAOA) on 3-regular graphs to obtain provable lower bounds on the approximation ratio for warm-started QAOA. Given an initialization angle θ, we consider warm-starts where the initial state is a product state where each qubit position is angle θ away from either the north or south pole of the Bloch sphere; of the two possible qubit positions the position of each qubit is decided by some classically obtained cut encoded as a bitstring b.

We illustrate through plots how the properties of b and the initialization angle θ influence the bound on the approximation ratios of warm-started QAOA. We consider various classical algorithms (and the cuts they produce which we use to generate the warm-start). Our results strongly suggest that there does not exist any choice of initialization angle that yields a (worst-case) approximation ratio that simultaneously beats standard QAOA and the classical algorithm used to create the warm-start.

Additionally, we show that at θ=60°, warm-started QAOA is able to (effectively) recover the cut used to generate the warm-start, thus suggesting that in practice, this value could be a promising starting angle to explore alternate solutions in a heuristic fashion.

4 - Exact Recovery Guarantees for Parameterized Non-linear System Identification Problem under Adversarial Attacks

Baturalp Yalcin, University of California, Berkeley, Berkeley, CA, United States, Haixiang Zhang, Javad Lavaei

In this work, we are interested in solving the system identification problem for parameterized non-linear systems using basis functions under adversarial attacks. Motivated by the LASSO-type estimators, we study the exact recovery property of a non-smooth estimator, which is a global solution to the embedded ℓ_1-loss minimization problem. First, we derive necessary and sufficient conditions for the well-specifiedness of the estimator and the uniqueness of global solutions to the underlying optimization problem. Next, we provide exact recovery guarantees for the estimator when the basis functions is bounded or Lipschitz. The non-asymptotic exact recovery is guaranteed with high probability, even when there are more noisy data than clean data. Finally, we numerically illustrate the validity of our theory. This is the first study for sample complexity analysis of a non-smooth estimator for non-linear system identification problem.

5 - New Sample Complexity Bounds for Convex Stochastic Programming Under Heavy-Tailed-Ness

Hongcheng Liu, University of Florida, Gainesville, FL, United States, Jindong Tong

We study the sample average approximation (SAA) and its simple regularized variation in solving convex or strongly convex stochastic programming problems. Under heavy-tailed assumptions and comparable regularity conditions as in the typical SAA literature, we show --- perhaps for the first time --- that the sample complexity can be completely free from any metric entropy that measures the complexity of the feasible region (e.g., logarithm of the covering number). As a result, our new bounds can be more advantageous than the state-of-the-art in terms of the dependence on the problem dimensionality.

TE35

Summit - 427

Network Modeling for Sustainable Mobility: Equity, Efficiency, and Multimodality

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Bingqing Liu, University of California, Los Angeles, Los Angeles, CA, United States

Co-Chair: Yuqiang Ning, University of Florida, Gainesville, FL, United States

1 - A Dynamic Coordinated Joint Routing and En-Route Charging Mechanism for Traffic Congestion Mitigation Built upon Distributed Asynchronous Optimization

Yuqiang Ning, University of Florida, Gainesville, FL, United States, Lili Du

As electric vehicles (EVs) grow in popularity, new challenges emerge regarding congestion at charging stations and on roads. Real-time traffic and charging station availability information enables travelers to make informed routing and charging decisions to avoid congestion. However, if travelers react independently to similar information, it can exacerbate congestion due to the flash crowd effect. This study aims to alleviate such problems by developing a dynamic coordinated joint routing and en-route charging mechanism (DcRC) to provide routing and charging guidance for a mix of EVs and internal combustion engine (ICE) vehicles, reducing congestion on roads and at charging stations. By considering traffic flow and charging station queueing dynamics, the DcRC is formulated as a mixed strategy congestion game with an equivalent mathematical programming model, generating equilibrium routing and charging decisions (ERCD) to mitigate the flash crowd effect and reduce traffic congestions without violating each vehicle's self-interest. To integrate the DcRC into online navigation services, an asynchronous distributed ADMM-aided Branch-and-Bound (DAB) solution algorithm is developed to efficiently solve the DcRC with hundreds of participants. The DAB decomposes the complex program into sub-problems using a customized branch and bound algorithm, and utilizes an asynchronous distributed ADMM to solve each sub-problem, leveraging individual vehicles' computing resources while sustaining resilience against vehicles' unstable computing and communication performance. Numerical experiments on the Sioux Falls network validate the DcRC's effectiveness in reducing congestions and system costs, and the DAB's efficiency in supporting real-time navigation services involving hundreds of participants with unstable performance.

2 - Evaluating Integrated Fixed-Route Transit and Microtransit System Designs with a Flexible Modeling Approach

Michael Hyland, University of California, Irvine, Irvine, CA, United States, Siwei Hu

Our research addresses a timely problem related to the design of integrated fixed-route transit and more flexible microtransit service. Specifically, we develop an agent-based simulation modeling framework to evaluate alternative integrated system designs. Unlike existing studies in the literature, our modeling approach can readily handle a wide variety of design variables. Moreover, our modeling framework provides a breadth of performance metrics of particular interest to transit agencies, including required subsidy per transit user and employment accessibility in addition to mode share, vehicle miles traveled, and user travel time and cost metrics. Our modeling framework includes (i) a logit mode choice model to capture endogenous demand for public transit, (ii) a supernetwork model with fixed-route transit, microtransit, and pedestrian infrastructure layers and a pathfinding algorithm to capture multi-modal path choice, and (iii) a mobility-on-demand fleet simulator, FleetPy, to capture the supply-demand dynamics of microtransit service. We illustrate the capabilities of the modeling framework by analyzing integrated system designs that vary the following design parameters: fixed-route frequencies and microtransit fleet size, service region structure, virtual stop coverage, and operating hours. We also analyze several different fare structures for microtransit. Preliminary results suggest that microtransit service increases system costs while substantially increasing accessibility to jobs, groceries, etc. Also, preliminary results suggest that partitioning one large microtransit service region into several smaller subregions may increase the number of fixed-route transit trips. Yet, it decreases overall accessibility, mobility, and non-car mode share.

3 - Promoting Sustainable Mobility: Behavioral Interventions for Transportation Sustainability and Equity

Viswa Sri Rupa Anne, Georgia Tech, Atlanta, GA, United States, Srinivas Peeta

Communities across the U.S. face transportation-related challenges of mobility, environmental sustainability, accessibility, and equity. Low adoption of sustainable modes such as transit and active transportation modes (e.g., walking, biking) along with high private vehicle usage negatively impact mobility and environmental sustainability by increasing congestion and greenhouse gas emissions. Personalized behavioral interventions such as informational messages, nudges, gamification, and tangible incentives (e.g., coins, points) that target different socio-demographic groups to promote sustainable routes or modes can represent a promising paradigm to address these challenges. Emphasizing the need for system-level improvements, this study proposes a behavioral intervention design. Unlike traditional travel behavior approaches that rely on penalty-based approaches such as tolls or congestion pricing, the proposed interventions are sustainable as they are not limited by funding, and equitable by being accessible to all income groups. A bi-level transportation network design problem is used to model and design behavioral interventions. The upper-level consists of a multi-objective optimization model with system-level mobility, environmental sustainability, accessibility, and equity as the objectives to design the interventions. The lower level formulates a user equilibrium model for travel behavior under the interventions. The study findings offer valuable insights for policymakers, urban planners, and transportation agencies seeking to promote sustainable mobility options. They contribute to advancing knowledge on the effectiveness of behavioral interventions in achieving broader societal goals related to transportation sustainability.

4 - Statistical Inference of Traffic States Using a Generalized Bathtub Model

Zihan WAN, The Hong Kong Polytechnic Univeristy, Hong Kong, Hong Kong, Wei Ma

The modeling of traffic dynamics in various scenarios is a key component in the management of intelligent transportation systems. The bathtub model, proposed in Vickrey (1991) and generalized in Jin (2020), incorporates the Macroscopic Fundamental Diagram (MFD) to model the traffic flow in a homogeneous region, in which the travel speed is solely determined by vehicle density. This framework enables the stochastic modeling of traffic dynamics. In real life, it is often the case in which we have access only to partial observations of traffic states, which are not enough for the decision makers to get a panoramic view of the traffic system. Therefore, based on the partially observed variables, an attempt to make inference for the underlying latent variables and other unknown parameters is valuable. In our work, we develop such a stochastic variational inference framework through a hierarchy of latent and observed variables, which is useful for stochastic demand estimation and parameter calibration. For example, suppose we have a time series consisting of the counts of trips that exit a region; based on speed-density relation induced by MFD, it is possible to infer the number of trips generated inside the region, the distribution of the trip distances, and MFD parameters, which are useful to improve the performance of the traffic system.

5 - Leveraging Partnership between Public Transit and Emerging Private Modes to Enhance Accessibility Equity for Low-Income Populations

Md Gulam Kibria, Georgia Institute of Technology, Atlanta, GA, United States, Srinivas Peeta

Low-income populations, largely consisting of transit-dependent individuals, often reside in transit deserts. They face significant challenges due to the first-mile/last-mile (FMLM) problem, which hinders their ability to access transit services and other essential needs such as employment, education, healthcare, and groceries. The landscape of urban mobility is undergoing a transformative shift with the emergence of private modes like ride-hailing and micromobility. This shift presents an opportunity for synergy between these emerging modes and public transit. By providing convenient FMLM connectivity, these emerging modes can complement transit, thereby enhancing access to transit services and other essential needs. This study explores how partnerships between public transit and emerging modes can be leveraged to address the accessibility issues faced by low-income populations. It develops an optimization model aimed at maximizing access to essential needs while minimizing accessibility disparities across different zones within a city through strategic allocation of transit lines and optimal fleet sizing for emerging transportation modes. Additionally, we propose an individual-level pricing mechanism designed to ensure affordability. It considers the unique sociodemographic characteristics of each individual to offer customized pricing structures tailored to varying income levels. By making transportation services more accessible and affordable for low-income populations, we aim to alleviate the cycle of poverty perpetuated by accessibility inequities. Our findings indicate that leveraged partnerships can significantly improve access for low-income zones and reduce accessibility disparities across different geographical areas. Moreover, our proposed pricing mechanism enhances both affordability and overall utility for each individual.

TE36

Summit - 428

Markets and Algorithms for Sustainable Transportation

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Manxi Wu, Cornell University, ORIE, ITHACA, NY, United States

1 - Network Flow Problems with Electric Vehicles

Haripriya Pulyassary, Cornell University, ORIE, Ithaca, NY, United States, Kostas Kollias, Aaron Schild, David Shmoys, Manxi Wu

Electric vehicle (EV) adoption in long-distance logistics faces challenges such as range anxiety and uneven distribution of charging stations. Two pivotal questions emerge: How can EVs be efficiently routed in a charging network considering range limits, charging speeds and prices? And, can the existing charging infrastructure sustain the increasing demand for EVs in long-distance logistics? This talk addresses these questions by introducing an EV network flow model that incorporates range constraints and nonlinear charging rates. We identify conditions under which polynomial-time solutions can be obtained for optimal single EV routing, maximum flow, and minimum-cost flow problems.

2 - Market Mechanism for Community-Based EV Charging

Minjae Jeon, Cornell, Ithaca, NY, United States, Lang Tong

In this talk, we will present our findings regarding the pricing mechanism within an energy community that includes electric vehicles. An energy community is a coalition of individual households possibly sharing their distributed energy resources, such as rooftop solar, energy storage, flexible controllable devices, and electric vehicles. We assume that this energy community is ran by a profit-neutral community operator who coordinates power transmission between the utility and the community members. The community operator incentivizes members through price signals, which they use to schedule their electric vehicles and other flexible devices, considering uncertainties in renewable generation.
We first formulated the individual scheduling problem as a stochastic dynamic programming problem with continuous state and action spaces. In general, solving such a problem is intractable but we have shown that the optimal scheduling policy follows a threshold structure based on the principle of procrastination, which delays charging EV until the last possible moment. Next, we investigated the pricing mechanism for

community operator. We formulated a community operator's central scheduling optimization, determining community energy price. We showed that these price incentives guarantee community operator breaks even. Additionally, individual community members are more profitable than staying outside of the community, following the price incentives provided by community operator.

3 - Equity and Efficiency for Congestion Pricing

Ben Rosenblad, University of Michigan, Ann Arbor, MI, United States

Advancing urban transit systems has emerged as a vital strategy for reducing vehicular congestion within city environments. Nevertheless, the underutilization of transit infrastructure is frequent, attributed to factors such as prolonged journey times and reduced comfort. To enhance transit ridership, implementing a blend of congestion pricing and transit subsidies (PS) is a widely endorsed policy. Under such a policy, within a bimodal transportation network that accounts for heterogeneous user valuation of time (VOT), at equilibrium, individuals with a VOT surpassing a set threshold invariably prefer to drive rather than utilize public transportation. Contemporary research has introduced an innovative "Credit Charge-Cum-Reward" (CCR) strategy, distinct in its capacity to sustain a state of equilibrium that fosters "mixed mode users"—commuters who interchange between transit and personal automobile use. This study conducts a comparative analysis of the two strategies, focusing on both efficiency and equity through the lens of the alpha-fair utility function, where the alpha parameter assigns the degree of emphasis on equity. Our findings suggest that in a system-optimal scenario prioritizing maximum efficiency, the phenomenon of mixed mode users is inviable. Moreover, under specific assumptions on the VOT distribution, the presence of mixed mode users is inviable, irrespective of the equity weight ascribed by alpha. Yet, we delineate VOT distribution characteristics where the existence of mixed-mode users can be conducive to enhanced equity, thus promoting a balanced and inclusive urban transportation system.

4 - Near-Optimal Mechanisms for Resource Allocation Without Money

Moise Blanchard, MIT, Cambridge, MA, United States, Patrick Jaillet

We consider a sequential resource allocation problem with strategic agents competing for a single resource at each step. Standard mechanism design approaches for resource allocation heavily rely on the use of money to align incentives. In some applications however, enforcing payments to the central planner may not be desirable. We focus on the setting when money transfers are not allowed. In this case, prior information on the agents is necessary and we suppose that the central planner has complete access to the utility distributions of the agents. Following the "promised utility" framework introduced in Balseiro et al. (2019), we first design non-monetary mechanisms for arbitrary number of agents and utility distributions, then show these are essentially optimal as the discount factor converges to one, or equivalently when the number of iterations T grows. In particular, we show that the average social welfare loss for the optimal mechanism is between 1/T and 1/T^(1/2), which corresponds respectively to smooth agent utility distributions and discrete distributions; and we can characterize the precise convergence rate for any input agent distributions.

TE37

Summit - 429

Sustainable Transportation Systems

Contributed Session

Chair: Lan Wang, Cal State University, East Bay, Hayward, CA, United States

1 - Go Green: Electric Vehicles in a Ride Hailing Platform

Xiuyi Zhang, Tsinghua University, Beijing, China, People's Republic of, Yongbo Xiao, Yue Zhang

Carbon emissions from ride-hailing are a serious concern for both ride-hailing platforms and governments. One way to address this concern is by allowing electric vehicle (EV) drivers to serve riders instead of fuel vehicle (FV) drivers. In this paper, we consider how a ride-hailing platform can introduce a green option that pairs riders exclusively with EV drivers. Specifically, we model three systems: the pooling system, where the wages for EV and FV drivers are equal; the semi-dedicated system, where the wages for EV and FV drivers are not equal; and the dedicated system, where the platform operates two separate subsystems for EV and FV drivers, and riders join one subsystem. We use the percentage of transactions fulfilled by EVs to measure the system's environmental performance and consider that the platform may face environmental constraints. We investigate the platform's decisions on prices and wages in each system and compare their performance in terms of profit and environmental impact. We show that, compared to the pooling system, the semi-dedicated system must be more environmentally friendly, but the dedicated system may not necessarily be so. We highlight the "mixing effect" of the semi-dedicated system and the "environmental concern effect" of the dedicated system and demonstrate that when riders' environmental concern is moderate, the dedicated system can achieve a "win-win" scenario.

2 - Online Operations of a Mixed-fleet Car-sharing Service System Catering Specialized Needs of the Senior

Xindi Tang, Central University of Finance and Economics, Beijing, China, People's Republic of, Xiaozhuang Li, Fang He, Yang Liu

In this study, we focus on the online operations of a mixed-fleet car-sharing service system. This system not only meets regular ad-hoc demands but also addresses the unique travel needs of seniors by integrating healthcare services. Given the increasing demand for personalized travel and an aging population, such integration is crucial. We propose a four-layer framework termed 'Chaining-Conducting-Matching-Relocation', designed to enable a closed-loop for dynamic operations management of the system. Additionally, we have developed algorithms to efficiently resolve online operational challenges. These algorithms approach problems through a learning-and-optimization method and successfully apply a model-supported implicit neural network with deep reinforcement learning for intelligent decision-making in car-sharing scenarios.

3 - Resilience and Vulnerability in Transportation Networks: A Comprehensive Review of Methods and Future Directions

Ahmad Ghasemkhani, The University of Oklahoma, Norman, OK, United States, Charles Nicholson

This research provides an overview of the current literature on the resilience of transportation networks. The role of these networks in promoting economic prosperity and societal well-being is crucial. With increasing challenges posed by natural disasters, there is a shift towards resilience-centric strategies that depart from conventional risk assessment methodologies. Studies examine how networks respond to disruptions, assessing factors such as recovery time, robustness, survivability, and the impact of aging and environmental aggressiveness on infrastructure performance. Vulnerability is often the primary focus when exploring resilience metrics, and maintenance plays a vital role in evaluating life-cycle resilience frameworks. Frameworks are evolving to accommodate network complexity, including multimodal aspects and interdependencies among diverse infrastructure systems. This research scrutinizes the role of graph theory in studying topological structures and the influence of recent advancements in data analysis and model simulation technology. It also emphasizes the importance of equitable resilience metrics and the need for a universally accepted measure of resilience. Also, this research addresses uncertainties and interdependencies, exploring approaches such as robust optimization and machine learning. Overall, the research provides a comprehensive overview of transportation network resilience analysis, highlighting key metrics, frameworks, and methodological advancements. It identifies research gaps and future directions to enhance the resilience of transportation networks.

4 - Optimal Subsidy and Pricing Decisions to Promote E-vehicle Adoption with Consideration of Range Anxiety and Market Segmentation

Lan Wang, Cal State University, East Bay, Hayward, CA, United States, Kunpeng Li

E-vehicles are a great solution to climate change by reducing carbon emissions. However, high upfront purchase cost and range anxiety are the top barriers against e-vehicle adoption. Purchasing subsidy, a government regulatory policy, has been an efficient tool in overcoming the cost barrier. In this paper, we study the roles of governments, firms and customers in promoting e-vehicle adoption for sustainable production and consumption, through subsidy, pricing and purchasing decisions respectively. We consider a market consisting of two segments of consumers who are heterogeneous in e-vehicle valuations. Both segments have range anxiety for e-vehicles. E-vehicle purchase decisions are influenced by consumers' range anxiety, preference for e-vehicles, social planner's subsidy level and retail prices for both e-vehicles and traditional vehicles. We formulate an analytical model with profit and social welfare maximization to study the impact of market segmentation, consumers' range anxiety and e-vehicles' environmental benefits on optimal subsidy and pricing decisions.

TE38

Summit - 430

AAS Best PhD Dissertation and Best Paper Award

Award Session

Aviation Applications

Chair: Luis Cadarso, Rey Juan Carlos University, Fuenlabrada, Spain

Co-Chair: Max Li, University of Michigan, ANN ARBOR, United States

TE39

Summit - 431

Optimization and Analysis in Complex Systems

Contributed Session

Chair: Bin Shi, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No.55, Zhong Guan Cun East Road, Beijing, 100190

1 - Multi-Echelon Mix and Blend Optimization in Grains Supply Chain

Madhushini Narayana Prasad, Cargill R&D, Houston, TX, United States

This work focuses on designing an agile supply chain model to optimize the flow of soybean and corn grains through a multi-echelon network of facilities located in and around Brazil. The key objective of this model is multifold such as, but not limited to, maximizing the throughput utilization of facility assets, maximizing the gains from mixing and blending and minimizing the demurrage of river barges, while proactively tackling business constraints such as limited storage capacity, drying capacity and higher production volumes during the peak harvest seasons. While the problem statement seems straight-forward, the main challenge arises from blending equations being quadratic in nature rendering the problem computationally intractable. We employ novel approximation techniques to replace the large MINLP formulation with a MIP model. In addition, multiple inherent conflicting objectives such as minimizing demurrage and maximizing asset utilization are handled via goal programming approach. Finally, the proposed bi-level MILP is solved using Gurobi, to derive a 15-day tactical plan of good solution quality within reasonable computational effort.

2 - Analyzing Panama Canal Transit Equilibrium: A Complex Queueing System with Priority Pricing and Auctions

Zhongjun Ma, Tsinghua University, Beijing, China, People's Republic of, Xiwen Bai, Xi Lin

The Panama Canal serves as a crucial gateway between the Atlantic and Pacific Oceans, facilitating approximately 3% of the world's maritime trade as of 2023. With a focus on business profitability and customer satisfaction, the Panama Canal Authority (ACP) has devised complicated navigation regulations to manage the growing demands for transits, employing market mechanisms such as priority pricing and slot auctions. This study delves into the equilibrium dynamics between customers and the ACP, modeling the Panama Canal as a complex queueing system. Specifically, we establish a novel queueing system model incorporating priority rules and heterogeneous customers to characterize the market-oriented transit policies of the Panama Canal and the diverse service restrictions imposed by vessels. The model captures the decision-making process of vessels across multiple stages: 1) vessels are afforded the choice to either join the queue or opt for an alternative route; 2) vessels may compete for a limited number of booking slots for priority transits; 3) winners of the slot auction process gain priority access. Equilibrium behaviors within the Panama Canal transit are analyzed based on the queueing system model from two dimensions, including individual maximization and social optimization. We investigate strategies of shipping companies to minimize their transit costs and total transit time. Additionally, we assess the impact of market mechanisms proposed by the APC on the profitability of the Panama Canal and the overall welfare of both shipping companies and the authority.

3 - Mapping the likely routes of illicit cocaine trafficking in Costa Rica

Hashir Tanveer, The University of Alabama, Tuscaloosa, AL, United States, Nicholas Magliocca, Nicholas Freeman

Globalization of economic networks provides opportunities to expand businesses not only in the legal domain but also in illegal trafficking, especially cocaine trafficking. Despite the global reach and economic scale of cocaine trafficking, our geographic understanding of the trade remains limited. Available information aggregated at the country level leads to assumptions of homogeneous activities across countries, while information at the actor level adds complexity to generalizing and scaling up to the transnational scope. A high-resolution and more spatially disaggregated understanding of how the cocaine supply chain embeds across multiple locations and the spatial movement of traffickers is necessary. Drug traffickers prioritize established routes for efficiency but alter them to avoid capture and deliver cocaine to maximize profit and meet demands from the U.S. and EU. Limited data exists on these routes, however, qualitative and quantitative data from media reports, drug-related seizures, airstrips, aggregation/storage locations, and likely geographical drop-off locations of cocaine from South America to Costa Rica's eastern Pacific shores provide insights into drug flow. This research aims to utilize these multisource data and route optimization modeling approach to map likely transportation routes used by traffickers to deliver cocaine from the eastern Pacific side to the Atlantic side of Costa Rica. The model employs various versions with different constraints and scenarios to incorporate stochasticity. Mapping these trafficking routes helps overcome spatial limitations and provides anticipation of trafficking network spread.

4 - Understanding the ADMM algorithm via High-Resolution Differential Equations

Bin Shi, Academy of Mathematics and Systems Science, CAS, Beijing, China, People's Republic of, Bowen Li

In the fields of statistics, machine learning, image science, and related areas, there is an increasing demand for decentralized collection or storage of large-scale datasets, as well as distributed solution methods. To tackle this challenge, the alternating direction method of multipliers (ADMM) has emerged as a widely used approach, particularly well-suited to distributed convex optimization. However, the iterative behavior of ADMM has not been well understood. In this paper, we employ dimensional analysis to derive a system of high-resolution ordinary differential equations (ODEs) for ADMM. This system captures an important characteristic of ADMM, called the \lambda-correction, which causes the trajectory of ADMM to deviate from the constrained hyperplane. To explore the convergence behavior of the system of high-resolution ODEs, we utilize Lyapunov analysis and extend our findings to the discrete ADMM algorithm. Through this analysis, we identify that the numerical error resulting from the implicit scheme is a crucial factor that affects the convergence rate and monotonicity in the discrete ADMM algorithm. In addition, we further discover that if one component of the objective function is assumed to be strongly convex, the iterative average of ADMM converges strongly with a rate O(1/N), where N is the number of iterations.

TE40

Summit - 432

Behavior Study in Supply Chain Management

Invited Session

Behavioral Operations Management

Chair: Xiaobo Zhao, Tsinghua University, Beijing

Co-Chair: Wanshan Zhu, Renmin University of China, Beijing, N/A

1 - Process Data and Lies in Supply Chain Collaboration

Fadong Chen, Zhejiang University, Hangzhou, China, People's Republic of, Zhi Zhu, Yingshuai Zhao

As remote communication via online platforms expands collaboration opportunities, it simultaneously reduces the opportunity cost of lying in information exchanges among potential supply chain partners. This study investigates the value of process data, response times (RTs) and mouse trajectories, in recognizing lies from truth-telling in online supply chain collaborations. Specifically, we consider a representative setting where a seller with private demand information sends a non-binding demand message to the manufacturer, who then decides the production quantity. Accordingly, the seller is motivated to inflate demand so that to ensure adequate supply. Our laboratory experiments demonstrate that RTs for lying are consistently longer and the Area Under the Curve (AUC) for lying is larger than those for truth-telling, unaffected by the detectability of lies or the collaboration duration. Moreover, the differences in RT and AUC between lying and truth-telling decisions increase with the seller's lying aversion. Employing machine learning, we further show that incorporating RT and mouse trajectory data significantly enhances the accuracy of distinguishing lies from truths. These findings underscore the effectiveness of process data in detecting lies in supply chain practice as well as the value of process data in capturing decision patterns for future research and management.

2 - Performance Improvement in Supply Chain Management: An Experimental Study

SHUAIKUN HOU, TSINGHUA UNIVERSITY, BEIJING, China, People's Republic of

Behavioral biases displayed by humans in decision-making can lead to either beneficial or detrimental effects on system performance. We propose a method aiming at mitigating the adverse effects of behavioral biases on system performance and promoting supply chain coordination. First, we conduct human-to-human experiments to model the behaviors of the supplier and the retailer under the buyback contract. We find that the retailer exhibits several behavioral biases, leading to non-coordinating contracts. Then we design bots to help implement buyback contracts in supply chain management: After a supplier proposes a wholesale price, the bots decide a buyback price according to coordination requirements. That is, the decision of bots makes that the game equilibrium of the supplier and the retailer in the behavioral models achieves supply chain coordination. The bots are called coordinating bots because of their purposes. We conduct human-to-human experiments of the buyback contract with coordinating bots, to understand the effect of coordinating bots on performance improvement. We find that the coordinating bots significantly improve the buyback contracts by always providing high-profit contracts thus leading to retailers ordering more. According to behavioral model analyses, the coordinating bots render the supply chain parties' profit allocation to be fairer under the buyback contract.

3 - Impact of Returns Risk in a Buyback Contract: An Experimental Study

Ao Zhuo, Renmin University of China, Beijing, China, People's Republic of, Wanshan Zhu

Manufacturers may fail to fulfill their buyback commitments within the buyback contract, thereby inducing returns risk for retailers. Furthermore, the probabilities of breaching these commitments vary among different manufacturers. Through laboratory experiment, our research aims to empirically test the impact of returns risk.

Our experimental study comprises two parts. In the first part, we conducted experiment with symmetric information of returns risk where both retailers and manufacturers are aware of buyback probabilities. The objective of this part is to examine the underlying impact of returns risk on supply chain decision-making. We observe significant mental accounting behavior among manufacturers and probability-weighting behavior among retailers.

In the second part, we focus on the scenario of asymmetric information, where manufacturers possess private information of buyback probabilities while retailers do not. The objective of this part is to test whether manufacturers indeed send signals separately based on their types, as hypothesized, and whether these signals are effectively received by retailers. Contrary to the theoretical prediction of separate equilibrium, we find that separating equilibrium is not favored when manufacturers set their contract prices. This suggests a divergence from traditional models and prompts an exploration of the underlying drivers. We posit the driver behind is mental accounting effect and we seek to uncover novel explanations for results.

TE41

Summit - 433

Graphs and Networks: Bridging Theory and Practical Applications in Causal Inference, Experimentation, Community Detection, and Ranking Aggregation

Invited Session

Applied Probability Society

Chair: Shuangning Li, University of Chicago, Chicago, IL, United States

1 - Spectral Ranking Inferences Based on General Multiway Comparisons

Mengxin Yu, University of Pennsylvania, Philadelphia, PA, United States

This paper studies the performance of the spectral method in the estimation and uncertainty quantification of the unobserved preference scores of compared entities in a general and more realistic setup. Specifically, the comparison graph consists of hyper-edges of possible heterogeneous sizes, and the number of comparisons can be as low as one for a given hyper-edge. Such a setting is pervasive in real applications, circumventing the need to specify the graph randomness and the restrictive homogeneous sampling assumption imposed in the commonly used Bradley-Terry-Luce (BTL) or Plackett-Luce (PL) models. Furthermore, in scenarios where the BTL or PL models are appropriate, we unravel the relationship between the spectral estimator and the Maximum Likelihood Estimator (MLE). We discover that a two-step spectral method, where we apply the optimal weighting estimated from the equal weighting vanilla spectral method, can achieve the same asymptotic efficiency as the MLE. Given the asymptotic distributions of the estimated preference scores, we also introduce a comprehensive framework to carry out both one-sample and two-sample ranking inferences, applicable to both fixed and random graph settings. It is noteworthy that this is the first time effective two-sample rank testing methods have been proposed. Finally, we substantiate our findings via comprehensive numerical simulations and subsequently apply our developed methodologies to perform statistical inferences for statistical journals and movie rankings.

2 - Spectral Clustering in the Gaussian Mixture Block Model

Shuangping Li, Stanford University, Stanford, CA, United States, Tselil Schramm

Gaussian mixture block models are distributions over graphs that strive to model modern networks: to generate a graph from such a model, we associate each vertex i with a latent feature vector u_i in R^d sampled from a mixture of Gaussians, and we add edge (i,j) if and only if the feature vectors are sufficiently similar, in that <u_i,u_j> > \tau for a pre-specified threshold \tau. The different components of the Gaussian mixture represent the fact that there may be different types of nodes with different distributions over features --- for example, in a social network each component represents the different attributes of a distinct community. Natural algorithmic tasks associated with these networks are embedding (recovering the latent feature vectors) and clustering (grouping nodes by their mixture component).
In this paper we initiate the study of clustering and embedding graphs sampled from high-dimensional Gaussian mixture block models, where the dimension of the latent feature vectors d goes to infinity as the size of the network n goes to infinity. This high-dimensional setting is most appropriate in the context of modern networks, in which we think of the latent feature space as being high-dimensional. We analyze the performance of canonical spectral clustering and embedding algorithms for such graphs in the case of 2-component spherical Gaussian mixtures, and begin to sketch out the information-computation landscape for clustering and embedding in these models.

3 - Experimentation for Different Scheduling Policies on Queueing Networks: Mixed Difference-in-Q Estimators Based on the Little's Law

NIAN SI, HKUST IEDA, Hong kong, Hong Kong, Nanshan Jia, Ramesh Johari, Zeyu Zheng

In data centers, tasks are dispatched to various servers to evenly distribute the workload. When a data center considers implementing a new scheduling algorithm, it typically conducts an A/B test prior to deployment to assess the real-world impact of this new method. However, a straightforward A/B test might be interfered with so-called ``Markovian'' interference. We utilized the difference-in-Q estimator, as developed by Farias et al. [2022], and introduced mixed Difference-in-Q estimators grounded in Little's Law. We show that our A/B testing methods significantly reduce bias and variance when testing various power-of-d policies. Extensive simulations were conducted under scenarios like non-stationary arrival rates, heterogeneous service rates, and communication delays. These simulations highlight the robustness and efficacy of our A/B testing approach.

4 - Design of Panel Experiments with Spatial and Temporal Interference

Tu Ni, Harvard Business School, Boston, MA, United States, Iavor Bojinov, Jinglong Zhao

Companies constantly test new features on millions of users, but a hidden issue can skew results: interference. This happens when someone's experience depends on what others get. Ignoring this can lead to misleading conclusions.

Researchers have proposed designs that limit interference, but they can be weak when testing periods are limited. This paper offers a novel design for "panel experiments" (testing across multiple groups over time) that accounts for both interference and lingering effects from past treatments.

Our design involves two key features: "random shaking," which groups users randomly, and "balanced temporal randomization," which randomizes treatment timing within groups. We prove this design's effectiveness mathematically, develop analysis techniques, and demonstrate its superiority through simulations, including real-world ride-sharing data. Companies can potentially achieve over 50% more reliable results with the same number of users by using this new approach.

TE42

Summit - 434

Flexibility Design and Demand Fulfillment for Service Platforms

Invited Session

Applied Probability Society

Chair: Yue Hu, Stanford University, Palo Alto, CA, United States

Co-Chair: Shixin Wang, The Chinese University of Hong Kong, Hong Kong, N/A

1 - Online Demand Fulfillment Problem with Initial Inventory Placement: A Regret Analysis

Irem Nur Keskin, Duke University, Durham, NC, United States, Alessandro Arlotto, Yehua Wei

We investigate a joint inventory placement and online fulfillment problem. At the beginning, the inventory of a single item is distributed to different warehouses. Then, at each period, an order arrives from one of the demand regions, and the decision maker makes an irrevocable decision on whether to accept or reject the order. In our model, we propose the minimum-inventory regret, a notion that includes both the selection of initial inventories and the performance of the selected fulfillment policy. We consider two state-of-the-art fulfillment policies: probabilistic fulfillment and score-based fulfillment. We prove that probabilistic fulfillment has a minimum-inventory regret that scales with the square root of the time horizon. On the other hand, we show that the score-based fulfillment policy has a minimum-inventory regret bound that is independent of the time horizon and polynomial with respect to the number of warehouses and demand regions. Our results have the following implication: the score-based fulfillment policy, when paired with offline inventory placement, outperforms probabilistic fulfillment with any inventory placement, and the performance gap increases with the time horizon.

2 - Flexibility designs for parallel server systems with general processing rates

Yuan Zhong, University of Chicago/Booth School of Business, Chicago, IL, United States, Gorkem Unlu

Process flexibility design is a central topic in operations management, but to date, all existing literature adopts a restrictive flow conservation assumption, which supposes that exactly one unit of resource capacity is required to process one unit of demand, no matter what the demand type is. We significantly relax this assumption by considering the design of sparse flexibility structures in parallel server systems with general processing rates. Our main contribution is a computationally efficient procedure that constructs, from any given system, a sparse flexibility structure with no more than m+n+min{m,n}-2 flexibility arcs, where m is the number of demand nodes and n the number of resources. We establish that in heavy traffic, our flexibility design has asymptotically optimal performance, as long as complete resource pooling can be achieved in the original system. This is joint work with Gorkem Unlu (Wayfair).

3 - Value of Sparse Structures in Dynamic Reusable Resource Allocation with Waiting

Yue Hu, Stanford University, Stanford, CA, United States, Jing Dong, Shixin Wang

We study the dynamic resource allocation problem in online service platforms featuring reusable resources, waiting space, and heterogeneous demands and resources. The service provider aims to balance the trade-off between maximizing revenue and minimizing waiting times across various demand types. For this problem, we propose a comprehensive framework to minimize the long-run average revenue loss and waiting cost by simultaneously designing the 1) flexibility structure, 2) admission control, and 3) scheduling policy. The flexibility structures are tailored to systems with varying levels of workload intensity, and the number of arcs in the network is linear in the number of resource and demand types. In these sparse networks, we show that simple static priority rules and threshold-based admission controls are asymptotically optimal in the many-server regimes. Furthermore, our proposed algorithm for designing system flexibility, along with the scheduling and admission control policies, is both easy to interpret and straightforward to implement. Numerical experiments demonstrate the effectiveness of our approach, particularly for smaller systems, in non-asymptotic environments.

4 - Online Rack Placement in Large-Scale Data Centers

Konstantina Mellou, Microsoft Research, Redmond, WA, United States, Saumil Baxi, Kayla Cummings, Alexandre Jacquillat, Rob McDonald, Ishai Menache, Marco Molinaro

This paper optimizes the configuration of large-scale data centers toward cost-effective, reliable and sustainable cloud supply chains. We formulate an integer optimization model that optimizes the placement of racks of servers within a data center to maximize demand coverage, adhere to space, power and cooling restrictions, and pace resource utilization for future demand. We also define a tractable single-sample online approximation (SSOA) approach to multi-stage stochastic optimization, which approximates unknown parameters with a single realization and re-optimizes decisions dynamically. Theoretical results provide strong performance guarantees of SSOA in the canonical online generalized assignment and online bin packing settings. Computational results using real-world data show that our optimization approach can enhance utilization and reduce power stranding in data centers. Following iterative improvements in collaboration with data center managers, our algorithm has been packaged into a software solution deployed in Microsoft's data centers worldwide. Deployment data indicate a significant increase in adoption, leading to improved power utilization, multi-million-dollar annual cost savings, and concomitant savings in greenhouse gas emissions. Ultimately, this paper constitutes one of the first large-scale deployments of a decision-making tool in data centers, contributing an interactive decision-making process at the human-machine interface.

TE43

Summit - 435

Drift Methods for Stochastic Systems

Invited Session

Applied Probability Society

Chair: Lucy Huo, Cornell University, Ithaca, NY, United States

Co-Chair: Jim Dai, Cornell University, Ithaca, NY, United States

1 - Explicit steady-state approximations for parallel server systems with heterogeneous servers

Yaosheng Xu, University of Chicago , Chicago, IL, United States, Jim Dai

The weighted-workload-task-allocation (WWTA) load-balancing policy is known to be throughput optimal for parallel-server systems with heterogeneous servers. This work concerns the heavy traffic approximation of steady-state performance for parallel-server systems operating under WWTA policy. Under a relaxed complete-resource-pooling condition, we prove that WWTA achieves a “strong form” of state-space collapse in heavy traffic and that the scaled workload for each server converges in distribution to an exponential random variable, whose parameter is explicitly given by system primitives. Various steady-state performance measures are shown to be approximated from this exponential random variable. Instead of proving a stochastic process limit followed by an interchange of limits – a method that dominates the literature, our method works directly with a pre-limit basic adjoint relationship (BAR) that characterizes the stationary distribution of each pre-limit system.

2 - Drift method in restless bandits

Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

Drift method in restless bandits

3 - Steady-State Convergence of the Continuous-Time JSQ System with General Distributions in Heavy Traffic

Jin Guang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Jim Dai, Xu Yaosheng

This paper studies the continuous-time join-the-shortest-queue (JSQ) system with general interarrival and service distributions. Under a much weaker assumption than the one in the literature, we prove that each station's scaled steady-state queue length weakly converges to an identical exponential random variable in heavy traffic. Specifically, we establish our results by only assuming $2+\delta$ moment on the arrival and service distributions for some $\delta>0$. Our proof exploits the Palm version of the basic adjoint relationship (BAR) approach as a key technique. Additionally, our method is applicable to the power-of-two choices.

4 - Uniform Moment Bounds for Multiclass Queueing Networks with SBP Service Policies in Multi-scale Heavy Traffic

Lucy Huo, Cornell University, Ithaca, NY, United States, Jim Dai

In Dai and Huo (2023), the authors justify that the stationary distribution of the scaled queue length vector process in multiclass queueing networks operating under static buffer priority (SBP) service policies has a product-form limit under a multi-scale heavy traffic condition. A major assumption is the moment state space collapse (SSC) property exhibited by the steady-state queue length process. In this work, we prove that the moment-SSC of scaled low-priority queue lengths holds if the moment-SSC of unscaled high-priority queue lengths holds, which has been carefully studied in the work of Cao, Dai, and Zhang (2022).

TE44

Summit - 436

Healthcare Innovation and Collaboration

Contributed Session

Chair: Alireza Kasaie, Loyola University Chicago, Forest Park, IL, 60130, United States

1 - The Assessment of the Emergency Medical Services Workflow for Implementing An AI-Driven Stroke Detection Tool to Improve the Patient Outcomes

Alireza Kasaie, Loyola University Chicago, Forest Park, IL, United States, Michael Saban, Samie Tootooni, Dima Elissa

Identifying severe stroke patients in Emergency Medical Services (EMS) is vital, as they need advanced care only available at comprehensive stroke centers. Current stroke scales, like the Cincinnati Prehospital Stroke Scale (CPSS), often lead to patients being taken to inadequate facilities, resulting in costly and time-consuming transfers. Our long-term goal is to implement an AI-driven tool in the EMS workflow. To ensure this solution meets existing needs, our study aimed to: (1) assess the inefficiencies of existing tools, (2) conduct customer discovery to identify key stakeholders and their needs, and (3) develop a workflow incorporating an AI-driven solution for EMS.

Using the I-Corps@NCATS Value Proposition Design framework, we conducted structured interviews with individuals across stroke centers and EMS systems in the Chicagoland EMS region. We collected EMS data from 2015-2020 to evaluate CPSS-based stroke assessments retrospectively.

The sensitivity and specificity for our region were 29.4% and 98.3%, respectively, significantly lower than the literature-reported 81.1% and 51.7%. This inefficiency was confirmed by several interviewees. We identified EMS medical directors as final decision-makers, EMTs as end-users, and other stakeholders, like dispatch agencies, as influencers. We provided an EMS workflow diagram and identified where our solution could be implemented with minimal changes to existing protocols, ensuring EMTs are not overburdened.

By validating the importance of accurate prehospital stroke diagnosis and assessing potential improvements in EMS workflow, our study significantly advances the development and implementation of AI-driven solutions in EMS. This study can serve as a valuable guideline for researchers focusing on similar solutions.

2 - Hospital collaboration network foster health resilience during pandemics

Lu Zhong, Rensselaer Polytechnic Institute, Troy, NY, United States, Jianxi Gao

Understanding hospital collaboration emerges as pivotal for ensuring healthcare resilience, especially during the COVID-19 crisis time. However, existing studies are limited by the dataset and method for quantifying such collaboration. This study addresses the gap by analyzing millions of patients' electronic health records across the United States. By mapping patient visits between hospitals, we build the collaboration network among hospitals in each state, and introduce a model to simulate the dynamics of hospital burden. Using this model, we identify the key strategies that facilitate efficient collaboration to mitigate the burden of disruption from pandemics. Supported by empirical evidence and analytical insights, this study highlights the critical importance of hospital collaboration in strengthening the resilience of healthcare systems.

3 - Optimal Integration of In-Person and Telemedicine Outpatient Consultations

Aparna Venkataraman, Department of Mechanical Engineering, Indian Institute of Technology Delhi, New Delhi, India, Varun Ramamohan, Sisira Edirippulige

In the evolving healthcare landscape, optimising the concurrent delivery of in-person and telemedicine outpatient consultations in hospital settings presents a complex operations management challenge. This study introduces a discrete-event simulation analysis to optimise a rolling-horizon patient appointment schedule for telemedicine and in-person consultations for both scheduled and walk-in patients. The study considers new and follow-up in-person and telemedicine consultations, including pre-scheduled appointments and walk-ins. Specifically, we consider the following operational questions: (a) Is having separate blocks for telemedicine consultations more efficient than an integrated in-person and telemedicine consultation schedule? (b) How many slots should be reserved for scheduled appointments versus walk-ins of each type (in-person versus telemedicine)? (c) What is the optimal consultation slot duration for each consultation type? (d) Corresponding to the optimal consultation slot duration, what should be the actual consultation slot duration to be communicated to patients? These questions are answered considering the varying levels of uncertainty in consultation durations, patient no-show rates and infrastructure disruptions for in-person and telemedicine consultations. The setting we consider is that of paediatric outpatient care in a large multi-speciality hospital in urban India, and the simulation is parameterised using real-world data from this setting. Through these experiments, we provide evidence-based recommendations for the optimal integration and scheduling of telemedicine and in-person consultations, underscoring the importance of data-driven decision-making in enhancing the efficiency of patient care in a modern healthcare delivery environment.

4 - Information Sharing in Outpatient Clinic Settings

Murray Côté, Texas A&M University, College Station, TX, United States, Jon Stauffer

A perennial challenge in outpatient clinic management is efficient patient flow. This hinges on patients to fully disclose the scope of their healthcare needs in advance. However, the full scope of a patient's health concerns may not be revealed and appointments are typically selected to satisfy a single primary health concern. We surveyed patients and providers from two large outpatient practices to identify the prevalence additional health concerns and patient and provider preferences in addressing additional health concerns. Our results provide important contributions to agenda setting in outpatient care and patient/provider communication. Additionally, information sharing is essential to better designed online appointment applications, improved patient flow, and provider utilization.

TE45

Summit - 437

Applications of Mathematical Models in Health and Public Policy

Invited Session

Health Applications Society

Chair: Suyanpeng Zhang, Northwestern University, Chicago, IL, United States

1 - Optimizing Vaccine Site Locations While Considering Travel Inconvenience and Public Health Outcomes

Suyanpeng Zhang, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen, Han Yu, Maged Dessouky, Fernando Ordonez

During the COVID-19 pandemic, there were over three million infections in Los Angeles County (LAC). To facilitate distribution when vaccines first became available, LAC set up six mega-sites for dispensing a large number of vaccines to the public. To understand if another choice of mega-site location would have improved accessibility and health outcomes, and to provide insight into future vaccine allocation problems, we propose a multi-objective mixed integer linear programming model that balances travel convenience, infection reduction, and equitable distribution. We provide a tractable objective formulation that effectively proxies real-world public health goals of reducing infections while considering travel inconvenience and equitable distribution of resources. Compared with the solution empirically used in

LAC in 2020, we recommend more dispersed mega-site locations that result in a 28% reduction in travel inconvenience and avert an additional 1,000 infections.

2 - Evaluating Co-Response Models for Crisis Calls

Veronica White, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Laura Albert

A crisis call is when one or more individuals suffer from a mental- or behavioral health-related issue that may require a police response. Many police departments have introduced specially trained crisis response teams to replace the traditional police response with new response paradigms that may involve sending different or multiple types of vehicles to a crisis call. We introduce queueing models and performance measures to capture the dynamics of new response paradigms. We evaluate the queueing models using discrete event simulation with a case study based on data from Seattle, WA and we compare the crisis response to a traditional response to elucidate performance across a range of input parameters.

3 - Improving the Equity and Efficiency of Cystic Fibrosis Screening

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States, Douglas Bish, Ebru Bish

Newborn screening detects life-threatening genetic disorders, one of the most prevalent is Cystic fibrosis (CF). More than 700 variants of the CFTR gene cause CF. These are detectable through genetic testing; CF is also associated with elevated immunoreactive trypsinogen (IRT) levels. For cost-effectiveness, most CF screening starts with an inexpensive biochemical IRT test; genetic testing is reserved for a small fraction of newborns with elevated IRT levels, based on a threshold. However, setting an optimal threshold presents challenges, particularly due to variations in IRT levels and CF prevalence among different racial and ethnic groups. The current practice of using a single IRT threshold for all newborns leads to inefficiencies and equity/fairness issues. Additionally, newborns' racial/ethnic identifiers are unreliable; and using race-based thresholds may lead to poor public perception.

We introduce an innovative screening approach, integrating an inexpensive, small-panel genetic test (DNA) before the IRT test, to develop a strategy that can customize IRT thresholds without relying on race-based information. This approach leads to novel optimization problems, involving variant selection for the small-panel DNA, and IRT threshold customization based on the number of variants detected by the small-panel DNA. (As an autosomal recessive disease, CF requires two variants.) We establish key structural properties of optimal IRT thresholds. The case study, utilizing CF data from the state of New York, compares the proposed approach with the current practice and race-based benchmarks, and demonstrates the potential improvements, in both the efficiency and equity of screening, possible by this novel approach.

4 - Metamodeling Colorectal Cancer for a User Interface and Improved Intervention Scenario Interpretation

Ashley Stanfield, North Carolina State University , Raleigh, NC, United States, Maria Mayorga, Meghan O'Leary, Kristen Lich

Colorectal cancer is the second most common cause of death by cancer in the United States. Additionally, many colorectal cancer deaths are preventable given increased screening rates. The NC-CRC Simulation model considers an input population, input screening levels, and the natural history of colorectal cancer to model colorectal cancer occurrence and progression. This model was used to create a user-interface, where policymakers could generate outcomes given the original input population. While this model is highly effective in predicting various outcomes such as cancer cases averted and life years lost, it is computationally expensive when multiple populations are considered. Our study uses metamodeling techniques to develop predictions for cancer cases and life years lost due to colorectal cancer. These metamodels are validated against results from the widely-used NC-CRC model. We explore diverse intervention scenarios, ensuring comprehensive coverage of potential policy options. Our approach includes the creation of individual-level metamodels, allowing for predictions for multiple input populations at no additional computational cost. These metamodels can be used to create a user interface that allows policymakers to input demographic information and generate predictions for cancer cases and life years lost at both individual and population levels. This conceptual framework facilitates informed decision-making by policymakers for colorectal cancer intervention planning.

5 - Single-Agent Learning Multi-Agent Decision Making (Slmd) & Ndash; a New Reinforcement Learning Approach for Epidemic Modeling

Sonza Singh, University of Massachusetts, Amherst, Amherst, MA, United States, Chaitra Gopalappa

Public health interventions during new epidemic outbreaks include simple measures, such as increasing awareness, and extreme measures, such as lockdowns, the trade-off being the costs and effectiveness. Taking jurisdiction-specific intervention measures are key to effective and optimal epidemic control. As such decisions are highly dynamic, reinforcement learning (RL) methods are suitable tools for identifying optimal decisions. Current RL algorithms include single-agent RL models and multi-agent RL (MARL) models. Single-agent RL models evaluate decisions in isolation to each jurisdiction, however, are insufficient representations of highly dynamic real-world epidemic settings. MARL models evaluate decisions specific to each jurisdiction while considering the interactions between all jurisdictions. They are computationally complex and this complexity grows with the number of jurisdictions, as each jurisdiction trains a separate deep learning model. This makes it challenging to solve large-scale problems, which are typical for epidemics. For example, in the United States, every county should have its own intervention policy based on the state of its epidemic. However, because of travel, there is a chance that infections from any county may spread to any other county, and thus these interactions should still be considered in the decisions made. To model the interactions between jurisdictions while reducing the computational complexity, we propose a new approach: single-agent learning multi-agent decision-making (SLMD). SLMD utilizes the unique features of epidemics to train a single deep learning model that considers the interactions between all agents, but each agent makes its own decision based on the state of their epidemic.

TE46

Summit - 438

Innovative Modeling to Inform Infectious Disease Prevention and Control

Invited Session

Health Applications Society

Chair: Yuming Sun, Georgia Institute of Technology, ATLANTA, GA, United States

Co-Chair: Lauren Steimle, Georgia Tech ISyE, Atlanta, GA, United States

1 - A Comprehensive Analysis of Methods to Incorporate Behavior in Agent-Based Models for Infectious Diseases

Sebastian Rodriguez Cartes, North Carolina State University, Raleigh, NC, United States, Maria Mayorga, Osman Ozaltin, Julie Swann

Individual behavior is the first protective barrier to prevent contagion from infectious diseases. Hence, accurate modeling of such actions is needed for simulation tools to represent the spread of a disease through a community. Starting from a theory-based modeling framework that incorporates people's beliefs and perceptions, we investigate the performance of several methods used to model individual-level behavior in agent-based models. We calibrated these approaches with survey data collected during the latest COVID-19 pandemic, and assess model performance in predicting future behavior. We aim to shed light on the usefulness of using empirical models to represent behavior. Moreover, we illustrate the impact of these models in quantifying differences in health outcomes between distinct demographic groups that are expected to behave differently. Our results highlight the need of incorporating individual behavior to properly evaluate equity goals in heterogeneous communities.

2 - Application of Discrete Event Simulation in Cost-Effectiveness Evaluation of Prognostic Tests: A Case of Host-Response-Based Transcriptional Signatures for Tuberculosis Among Migrants to the United States

Yuli Lily Hsieh, Interfaculty Initiatives in Health Policy, Harvard University, Boston, MA, United States, C Robert Horsburgh Jr, Ted Cohen, Jeffrey W Miller, Joshua A Salomon, Nicolas A Menzies

Purpose

Host-response-based transcriptional signature (HrTS) tests have been developed to identify individuals at high near-term risk of progressing to tuberculosis (TB) disease (“incipient” TB). The accuracy of these prognostic tests depends on the time between testing and TB onset, while individuals’ times-to-TB depend on individual risk factors. In this study, we show that discrete event simulation can be useful for evaluating the cost-effectiveness of prognostic tests in post-arrival screening strategies among migrants to the U.S.

Methods

We used a discrete-event simulation model to compare four strategies: (1) no screening for TB infection or incipient TB; (2) ‘IGRA-only’, screen all with interferon gamma release assay (IGRA), provide TB preventive treatment for IGRA-positives; (3) ‘IGRA-HrTS’, screen all with IGRA followed by HrTS for IGRA-positives, provide incipient TB treatment for individuals testing positive with both tests; and (4) ‘HrTS-only’, screen all with HrTS, provide incipient TB treatment for HrTS-positives. We assessed outcomes over the lifetime of migrants entering the U.S. in 2019 and conducted sensitivity analyses to evaluate the robustness of results.

Results

The IGRA-only strategy dominated the HrTS-based strategies under both healthcare sector and societal perspectives, with incremental cost-effectiveness ratios of $78,943 and $89,431 per quality-adjusted life-years (QALY) gained, respectively. This conclusion was robust to varying costs ($15–300) and characteristics of HrTS, and the willingness-to-pay threshold ($30,000–150,000/ QALY gained), but sensitive to the rate of decline in TB progression risk after U.S. entry.

Conclusions

Discrete event simulation can be a useful decision analytic modelling approach for evaluating prognostic tests.

3 - Optimizing Outbreak Response: Using a Spatiotemporal Model of Disease Spread to Support Polio Eradication

Arie Voorman, Gates Foundation, Seattle, WA, United States

One of the key challenges for the Global Polio Eradication Initiative is to prevent outbreaks from spreading to new areas by containing and stopping them quickly, typically through house-to-house vaccination campaigns. Outbreak response campaigns must cover all infected populations, but are often constrained by available budget and vaccine supply. Further, the sparse and delayed detection of poliovirus make it hard to determine the size of the infected area and the trend of an outbreak in real-time.

We introduce a spatiotemporal statistical model that we use to support outbreak response activities and discuss how we collaborate with decision makers in challenging operational contexts in Africa and the Middle East. We provide an overview of the technical performance in terms of predictive value, the strategic implications of the model, and how it is integrated into the decision-making process of the Global Polio Eradication Initiative.

4 - Mechanistic Modeling of Social Conditions into Disease Predictions for Public Health Intervention-Analyses: Application to Hiv

Amir Khosheghbal, University of Massachusetts, Amherst, Amherst, MA, United States, Chaitra Gopalappa

Epidemic models typically simulate the spread of diseases as functions of behaviors, e.g., sexual and care behaviors for sexually transmitted diseases. However, multi-level factors, including poverty, housing or food insecurity, mental health, and substance use disorder etc., (we will collectively refer to these as social determinants of HIV (SDH)), are drivers of those behaviors. Therefore, there is increasing awareness for the need to model SDH into epidemic simulations to evaluate structural interventions alongside behavioral and care interventions.

However, the multivariate joint associations between SDH and behaviors are not available. Data for SDH are mostly available as county level marginal distributions, and associations between SDH and behaviors are mostly bivariate.

We combined copula probability theory and probabilistic graphical models for estimating the multivariate joint distributions. We estimate bivariate associations between SDH using copula. As data for associations between SDH and behaviors are mostly available through small sample literature studies, we used the bivariate SDH associations derived using copula and between SDH and behaviors from literature studies, as links in undirected graphical models, to estimate the joint density.

To illustrate, we used the joint distributions to model HIV-risk related behaviors as function of SDH in a national-level HIV/AIDS (PATH 4.0) model and studied the impact of hypothetical 100% efficacious SDH interventions on HIV prevention. We found that this intervention could lead to a cumulative 10-year reduction of 29% in HIV incidence.

TE47

Summit - 439

Healthcare Analytics

Invited Session

Health Applications Society

Chair: Susan Lu, Purdue University, West Lafayette, IN, United States

Co-Chair: Yao Li, Southern University of Science and Technology, N/A

1 - Inter-Organizational Healthcare Delivery between Emergency Departments: A Dynamic Network Perspective

Yao Li, Southern University of Science and Technology, Shenzhen, China, People's Republic of, Wenwen Li, Yukun Yang

To explore the interorganzational care delivery with tangled structure, we empirically investigate the ED-to-ED interhospital transfer from a dynamic network perspective, which is little addressed by existing studies. Using interhospital ED-to-ED transfers from 178 hospitals in the State of Florida from 2013 to 2017, our temporal exponential random graph model (TERGM) analysis provides encouraging empirical evidence.

2 - Combatting Opioid Use Disorder: The Role of Drug Dosage Alert Systems

Lin Qiu, Southern University of Science and Technology, ShenZhen, China, People's Republic of, Yongchun Fang, Yao Li

The opioid epidemic has swept the United States. Opioid overprescribing by prescribers is a primary contributing factor, as it could facilitate opioid use disorder. Drug dosage alert systems provide real-time alerts to prescribers when they prescribe doses exceeding the recommended ranges. This would prevent opioid overprescribing and consequently reduce opioid use disorder. Utilizing a hospital-level panel dataset encompassing 165 hospitals in New York State from February 2012 to December 2017, we employ difference-in-differences and counterfactual estimators to analyze the impacts of drug dosage alert systems implementation. Our results suggest that the implementation of these systems significantly reduces outpatient hospital visits due to opioid use disorder. Additional analyses suggest that the effectiveness of these systems is moderated by the number of full-time physicians and dentists within hospitals. Through an analysis of panel data involving 231 prescribers and 16,868 observations, we find that a reduction in the dosages of opioid prescriptions following system adoption may explain the observed effectiveness. Specifically, drug dosage alert systems significantly reduce the prescribed dosage of opioid medications. This effect is moderated by the abuse potential of the prescribed opioids and the prescribers' years of practice. Our findings highlight the important role of these systems in promoting safer prescribing practices and curbing opioid use disorder.

3 - Evolution of Patient Perceptions and Experiences with Telehealth: Insights from Reddit Communities

Xinying Liu, University of Houston - Clear Lake, Houston, TX, United States

Telehealth has evolved over multiple decades in the US, with the COVID-19 pandemic acting as a catalyst for the removal of regulatory and reimbursement barriers, thereby accelerating its adoption. Despite this rapid growth, there remains a significant gap in understanding patients' perceptions of telehealth, which is crucial for optimizing healthcare delivery and ensuring patient satisfaction.

In this research, we investigate the evolution of patient perceptions toward telehealth over time and identify the factors influencing these changes through text mining. Additionally, we focus on patients with mental disorders, specifically those with anxiety, to understand how telehealth affects their healthcare experiences.

Data is collected from Reddit, a highly popular social media platform where user posts are organized by subject into user-created boards called "subreddits." We examine three subreddits: r/telemedicine, r/telehealth, and r/anxiety.

This research provides valuable insights into the dynamic nature of patients' perceptions of telehealth. Understanding how these perceptions change over time can inform strategies to enhance telehealth adoption, address barriers, and tailor services to meet evolving patient needs. By examining telehealth utilization, satisfaction, and preferences, healthcare providers and policymakers can better design interventions to sustain positive patient experiences and increase the acceptance of telehealth services.

4 - Is Continuity All We Need? A Modeling Approach to Evaluating Relational Continuity in Primary Care

Nicos Savva, London Business School, London, United Kingdom, Naireet Ghosh, Yueyang Zhong

Relational continuity of care, defined as a consistent therapeutic relationship between a patient and a physician across multiple healthcare events, has been shown to improve health outcomes, thereby reducing the load on the healthcare system through fewer physician visits.

Despite these benefits, maintaining continuity presents significant operational challenges by restricting resource pooling among physicians, leading to longer patient wait times and potentially increasing system load. Our study employs a novel queueing model to analyze the tradeoffs involved and to explore patient decision-making strategies between waiting for their regular physician or opting for another with shorter wait times. Our analysis reveals that, from the perspective of a social planner, it is optimal to maintain intermediate levels of continuity wherein the patients with the most complex needs visit their regular physician while those with lower complexity visit the physician that is available earliest. This balance depends on factors such as the size of the primary care practice (PCP), the patient mix, and the level of congestion. If, on the other hand, the continuity decision is left to individual patients, they are unable to account for the externality of their decision and, consequently, end up in an equilibrium where all patients choose to either always maintain continuity or always abandon it. Our study identifies conditions under which such suboptimalities are most pronounced, offering policymakers insights to guide decisions on the desired level of relational continuity, ensuring better patient outcomes and more efficient system functioning.

TE48

Summit - 440

Public Health Analytics and Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Retsef Levi, MIT, Cambridge, MA, United States

Co-Chair: El Ghali Zerhouni, MIT, Cambridge, United States

1 - Rapid Response Teams for Proactive Sepsis Treatment

Chia-Hao Chang, Columbia University, New York, NY, United States, Vineet Goyal, Carri Chan, Benjamin Ranard

Sepsis is an emergent medical condition in which one's immunological responses cause end-stage organ dysfunction and death. In a typical year, around three hundred and fifty thousand people die of sepsis in the United States. Its high mortality renders the timeliness of detection and treatment crucial. Some recent evidence has shown that providing early treatment to patients before they develop sepsis may lower their mortality. Often, early treatment for sepsis is administered by a rapid response team (RRT), which is a team of clinicians. However, the increased workload resulting from proactive care may block the emergent needs of other patients, including those who already have sepsis. Hence, the focus of this work is to design a policy to manage RRT for proactive sepsis treatment in a resource constrained setting to improve overall patient outcomes. Motivated by the structural insights from fluid approximation, we propose a one-step-look-ahead reservation policy. We justify our policy by showing its long-run asymptotic optimality. Finally, we perform simulations using the data from Columbia University Irving Medical Center//NewYork Presbyterian Hospital. Our simulations demonstrate that our proposed policy can potentially save an additional life following a shock of patient admissions to the hospital.

2 - Multi-Wave Infections: The Role of Variants and Population Immunity

Retsef Levi, Massachusetts Institute of Technology, Cambridge, MA, United States, Yanzhe Ma, El Ghali Ahmed Zerhouni

Respiratory viruses are characterized by their infection spread into multiple successive waves. To explain these patterns, biological and epidemiological evidence suggest that the emergence of new variants with immune-escape mutations can enable them to target previously immune sub-populations and lead to new waves of infection. Yet, existing epidemiology models, such as the Susceptible-Infectious-Recovered (SIR) model and its derivatives, do not yet account for these dimensions. Although these models enable to mathematically obtain the multi-wave infection spread patterns, they still rely on artificial modifications of their parameters during the infection spread process. This talk introduces the *Immunity-Variants Epidemic (IV-Epidemic)* model to capture the observed biological and epidemiological drivers of multi-wave infections into a single mechanism. By leveraging SARS-CoV-2 data to calibrate the model parameters, the study shows that the *IV-Epidemic* model can replicate the observed multi-wave infection spread patterns standalone. It also reproduces similar distributions of infections among variants and the observation that every new infection wave is driven by a few new variants. Additionally, it enables to assess the effect of pre-existing immunity on the early infection spread. The study thus provides several useful implications to better anticipate the spread of a new respiratory virus and enhance preparedness for future pandemics.

3 - Harmonizing Safety and Speed: A Human-Algorithm Approach to Enhance the FDA&Rsquo;S Medical Device Clearance Policy

Mohammad Zhalechian, Indiana University, Kelley School of Business, Bloomington, IN, United States, Soroush Saghafian, Omar Robles

For a medical device to be approved under the FDA’s Premarket Notification 510(K) pathway, the manufacturer can submit a claim indicating that it is as safe and effective as another legally marketed device. This vague regulatory process has led to a high recall rate for devices approved under this pathway, raising important concerns over the approach used by FDA. We develop a combined human-algorithm methodology to assist the FDA in improving the 510(k) medical device clearance process by reducing the risk of potential recalls and the workload of the FDA. We first develop machine learning methods to estimate the risk of recall of 510(k) medical devices based on the information available at the time of submission. We then propose a data-driven clearance policy that recommends acceptance, rejection, or deferral to FDA's committees for in-depth evaluation. We conduct an empirical study using a unique large-scale dataset, and our results show a 38.9% improvement in the recall rate and a 43.0% reduction in the FDA's workload. Our analyses also indicate that implementing our policy could result in significant cost-savings.

4 - Enhancing Uptake and Efficiency in Healthcare: Behavioral Interventions and Group Consultations

Soledad Giardili, University of Edinburgh, Edinburgh, United Kingdom, Kamalini Ramdas, Monica Heller, Amalia Miller, Sanjay Jain

Traditional clinician-patient consultations often involve repetitive exchanges of information and advice. A proposed solution to alleviate workload pressures in healthcare is group consultation, which offers a structured yet socially interactive environment for managing health in a group setting. As a relatively new concept in primary care, uptake might initially be low. We conducted a randomised controlled trial among

4,539 women aged 45 to 60 years to test the effect of behavioural interventions on the uptake of group consultations for menopause management. We introduced two intervention arms: peer testimony, where individuals with prior experience in group consultations shared their experiences, and expert description, which provided more detailed insights from healthcare professionals about the group consultation process. The peer testimonial intervention led to a 15.3 percentage point increase in group clinic uptake, while the expert description intervention resulted in a 13.4 percentage point increase, compared to an 11.1% baseline in the control group, who received the standard invitation. Additionally, both interventions significantly enhanced decision-making variables, such as certainty and clarity about the benefits of group consultations. Lastly, among participants who initially preferred a one-on-one appointment or no appointment, 21.5% changed their preference to a group consultation when it was offered at an earlier date. This shift underscores that patients are sensitive to the scheduling of their appointments. Our findings inform the design of behavioural nudges for promoting group consultations among the population and demonstrate how operational efficiency, achieved by shortening waiting times, significantly influences participants' decision-making processes.

5 - MedDRA-Standardized Drug Label Change Dataset for Postmarketing Analysis

Bonnie Akhavan, MIT, Cambridge, MA, United States, Retsef Levi

Adverse drug events (ADEs), otherwise informally known as drug side effects, are a widespread concern with detrimental consequences for patients and healthcare systems alike. We present a comprehensive drug label change dataset of over 45,000 entries which captures the listed adverse events for more than 5,000 prescription products. Significant label changes are extracted from a chronological label version history which we compile for each available product. After aggregating label changes, we utilize semi-supervised learning techniques to classify each entry into one of 27 standardized MedDRA System Organ Classes (SOCs). Through this classification, we enable the comparison of drug adverse event profiles across different products despite medication labels not adhering to a standardized medical terminology. We assess the accuracy of our SOC classification model on a labeled validation set compiled in collaboration with pharmacovigilance experts. To our knowledge, this is the first public dataset of its kind with comprehensive, historical drug label change information and associated MedDRA SOC classifications. By developing this dataset, we hope to facilitate a better understanding of ADE trends across drug classes in the postmarketing setting. Additionally, this dataset has potential to enable predictive technologies which could be used to detect unlisted ADEs earlier. In turn, early detection of these side effects could lead to improved clinical trial design and better patient outcomes.

TE49

Summit - 441

New Topics in Dynamic Assignment

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Parshan Pakiman, Chicago Booth, 5751 S Woodlawn Ave, Chicago, IL, 60637, United States

Co-Chair: Daniel Adelman, University of Chicago, Booth School of Business, Chicago, IL, United States

1 - Dynamic Assignment of Personalized Coupons under Reference Price Effects

Will Ma, Columbia University, Cambridge, MA, United States, Jackie Baek, Dmitry Mitrofanov

This work is motivated by our collaboration with an online retailer with a large email marketing campaign, where millions of customers each receive a daily email with a personalized discount amount. Based on past data, we build a personalized reference price model for each customer, where they are most likely to purchase if they receive a large discount after receiving small discounts for many days. There is a monthly promotional budget constraining the total discount that can be claimed across all customers. We introduce a new personalized reference price model and derive optimal controls for this model for long-term revenue, based on cycling different coupon values. Our analysis also simplifies understanding of reference price models, which could be of independent interest.

2 - Optimal Sharing of a Reusable Medical Resource During a Pandemic

Canan Uckun, University of Illinois at Chicago, Chicago, IL, United States, Daniel Adelman

Motivated by ventilator sharing between states in the US during the recent COVID-19 pandemic, we develop an optimal control model that maximizes lives saved by transferring ventilators across geographic regions to take advantage of inter-temporal differences in peak times.Our model captures the patient holding times of ventilators, which in practice was on the order of weeks, and show that ignoring them results in significant error in the number of ventilators needed to save a fixed number of lives, and vice versa. More generally, our framework could be used in the next pandemic to optimize the flow of any reusable medical resource to save lives.

3 - Optimizing Nurse-Shift Assignments Considering Nurse Preferences with No-Shows and Overbooking

HEDIEH SAZVAR, Northwestern University, Evanston, IL, United States, Seyed Iravani, Sanjay Mehrotra

The escalating demand for nurses, driven by an aging population and increased preventive care, has exacerbated the nursing shortage crisis and left hospitals struggling to fill shifts. This paper addresses the challenge of optimizing nurse-shift assignments to maximize the number of vacancies filled. Our model considers the different preferences and show-up probabilities for each nurse-shift pair, as well as shifts with different vacancies.

We propose a nurse-shift assignment model for implementation on a matching platform. The goal is to maximize the number of vacancies filled by nurses by ensuring that they choose to work the shifts and show up as scheduled, using a static policy. Given the infeasibility of directly solving the stochastic optimization problem due to its large state space, we characterize the optimal policy within a static framework to effectively manage large-scale scenarios. A heuristic solution method is proposed to closely approximate the optimal policy, and an overbooking strategy is integrated to mitigate potential losses from no-shows.

4 - Dynamic Assignment of Jobs to Workers with Learning Curves

Parshan Pakiman, University of Chicago Booth School of Business, Chicago, IL, United States, Daniel Adelman, Adam Mersereau

We study the problem of dynamically assigning jobs to workers over an infinite time horizon where the availability of both jobs and workers evolves stochastically over time. Workers develop familiarity with jobs when assigned to them, with higher familiarity levels resulting in lower costs when workers are assigned to the same jobs in the future. This problem gives rise to a high-dimensional Markov decision process (MDP) with stochastic combinatorial constraints encoding feasible job-worker assignments and endogenous-state dynamics capturing the evolution of familiarity levels. We propose a class of tractable control policies that relax the combinatorial constraints, another class of policies based on ignoring state dynamics, and generalizations of both classes. We provide supporting theory and insights into the policies' performances and numerically test them on different problem instances.

TE50

Summit - 442

Empirical Perspectives on Healthcare Challenges

Invited Session

MSOM: Healthcare

Chair: Michael Freeman, INSEAD, Singapore, Singapore

1 - From Full-Time to Flexi: Investigating the Impact of Workforce Shifts in Primary Care on Service Provision and Quality

Michael Freeman, INSEAD, Singapore, Singapore

Primary care practices worldwide are experiencing a shift away from traditional full-time roles toward increased part-time and temporary staffing models. This transformation necessitates understanding the implications for primary healthcare delivery. Our paper provides an empirical analysis examining the evolving work patterns and schedules of general practitioners (GPs) in the UK over the past decade. Using primary care records, we first examine how GP work volume impacts healthcare service utilization and patient outcomes. We also explore factors that may mediate this relationship. Going further, given the seeming inevitability of more flexible staffing models, our investigation also aims to inform effective workforce strategies that optimise healthcare delivery. Specifically, we examine scheduling approaches that may allow primary care practices to adapt to part-time and locum GPs. Our findings aim to inform healthcare leaders seeking to adapt to emerging flexible staffing models while maintaining high-quality and accessible primary care.

2 - To Consult or not to Consult - When Should Algorithms Seek Physicians' Advice for Predicting Surgery Durations?

Laura Poreschack, University of Cologne, Cologne, Germany, Dominik Walzner, Andreas Fuegener, Ulrich Thonemann, Christof Denz

Humans are a scarce resource, which is particularly problematic in healthcare. Hospital understaffing can have negative effects on the employees' physical and mental health as well as the quality of care. Thus, automation with AI is a valuable strategy to reduce human effort. In many healthcare management applications, AI augmented by human expertise has proven to perform better than full automation or humans working alone. However, while augmentation might increase the performance, it still requires human input, and thus does not decrease the human effort, which would be the case if humans were replaced by the AI.

In this paper, we propose a new approach to human-AI collaboration in healthcare operations addressing both performance and human effort. We develop a framework to decide which tasks should be automated and for which tasks the algorithm should seek input by a physician. We evaluate our framework considering the case of surgery duration predictions based on a data set from a large university hospital. The data set comprises 70,610 surgeries with 3,163 different procedure types. For each surgery, we have information on specific characteristics that can be grouped into patient-related, surgery-related, and physician-related features.

We observe that while human-augmented AI leads to an overall better performance than full automation, employing our framework can lead to an even higher forecasting accuracy with considerably lower human effort. Our results indicate that a partial automation of tasks that are easy for the AI allows physicians to focus on their medical duties instead of administrative tasks.

3 - Female FDA Inspectors and Medical Device Plant Inspections

Shirin Shahsavand, Washington State University, Pullman, WA, United States, Kevin Mayo, George Ball, Kaitlin Wowak

The Food and Drug Administration inspects medical device manufacturing plants on a recurring basis to ensure compliance with federally-regulated quality standards. FDA inspectors should apply these standards in an objective manner regardless of who inspects the plant. Research that examines how males and females perform their work differently, however, indicates this may be a flawed expectation. We examine how the gender of the FDA inspector influences inspection outcomes and the subsequent quality of medical devices manufactured at the plant. We theorize that female FDA inspectors will assign the most serious inspection outcome, an Official Action Indicated (OAI), more frequently than male inspectors. We also propose that OAIs assigned by female inspectors will lead to higher quality products made at the plant than OAIs assigned by male inspectors. Analyzing 19,722 inspections conducted at 4,590 plants in the U.S. from 2005 to 2019, we find that female inspectors assign significantly more OAIs than male inspectors. In addition, OAIs assigned by female inspectors lead to higher quality products from the plant compared to OAIs assigned by male inspectors. An instrumental variables analysis in robustness checks indicates a low risk of endogeneity bias in our model. Post-hoc analysis demonstrates support for a mechanism explaining our results, female inspectors take more time in conducting inspections that lead to OAIs than male inspectors. We offer contributions to research on product quality, plant inspections, and gender diversity, as well as implications for the FDA about how inspector gender may impact regulatory oversight and medical device quality.

4 - Opioid Scarcity and Health Outcomes

Basil Isaac, University of Michigan, Ann Arbor, MI, United States, Xinyu Liang

This paper examines the effects of a 2017 shortage of parenteral opioids, leveraging geographic variation to provide insights into how such shortages exacerbate care delays and opioid misuse.

Morphine, Fentanyl, and Hydromorphone are parenteral opioids given to patients who require fast and reliable analgesia on being admitted, and to those who have undergone surgeries. They are subject to DEA's production quotas aimed at containing illegal diversion of drugs. Therefore, their shortages are caused by both manufacturing issues and policy levers exercised by the DEA. The 2017 shortage of parenteral opioids was precipitated by supply side issues for the biggest manufacturer of the opioids which was compounded by supply restrictions imposed by the DEA.

We use two sources of data. First, the DEA's administrative data on legal shipments of opioids. This dataset, called ARCOS, was made publicly available for the period between 2006 and 2019 because of a court order. The second source is the Merative Truven Databases, which contain outpatient, emergency department (ED), and inpatient services, and drug prescription usage.

In this paper, we use a geographically varying measure of the incidence of the parenteral opioid shortage. We follow two different approaches to address the potential endogeneity between the volume of shipments and demand-side factors. First, we construct a shift-share instrument which exploits features of this specific shortage. The second approach is an event-study design which categorizes states as control and treatment states.

TE51

Summit - 443

Fairness and Matching in Service Operations

Invited Session

MSOM: Service Operations

Chair: Luyi Yang, University of California, Berkeley, Berkeley, CA, United States

Co-Chair: Shiliang Cui, Georgetown University, McDonough School of Business, McLean, VA, United States

1 - Treating to the Priority in Heart Transplantation

Philipp Afeche, University of Toronto, Rotman School of Management, Toronto, ON, Canada, Burhan Sandıkçı, Sait Tunc

The US heart transplantation system prioritizes transplant candidates according to the severity of their pre-transplant medical therapy: *The more severe the therapy, the higher the priority*. It is widely suggested that this rule opens up room for gaming the system, by over-treating certain patients to increase their priority on the waiting list. To study this issue, we consider an overloaded multi-class priority queueing model in a game-theoretic framework that captures the interplay among the key factors that may affect a transplant center's gaming behavior. The analysis of this model yields results on the conditions that yield such strategic gaming, and how it this behavior can be mitigated.

2 - Tight Bounds for the Price of Fairness

Daniel Granot, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada, Yifeng Cao, Yichuan Ding

A central decision maker (CDM), who seeks an efficient allocation of scarce resources among a finite number of players, often has to incorporate fairness criteria to avoid unfair outcomes. Indeed, the Price of Fairness (POF), a term coined in Bertsimas et al. (2011), refers to the efficiency loss due to the incorporation of fairness criteria into the allocation method. Quantifying the POF would help the CDM strike an appropriate balance between efficiency and fairness. In this paper we improve upon existing results in the literature, by providing tight bounds for the POF for the proportional fairness criterion for any n, when the maximum achievable utilities of the players are equal or are not equal. Further, while Bertsimas et al. (2011) have already derived a tight bound for the max-min fairness criterion for the case that all players have equal maximum achievable utilities, we also provide a tight bound in scenarios where these utilities are not equal. Finally, we investigate the sensitivity of our bounds and the bounds of Bertsimas et al. (2011) for the POF to the variability of the maximum achievable utilities.

3 - Not All Lines are Skipped Equally: An Experimental Investigation of Line-Sitting and Express Lines

Sezer Ulku, Georgetown University, Washington, DC, United States, Abdullah Althenayyan, Shiliang Cui, Luyi Yang

In this paper, we investigate how line-sitting and express lines affect customers' satisfaction and fairness perception about queues. In three experiments, we show that customers who encounter a line-sitting instance report higher satisfaction with their overall queueing experience than those who encounter an express-line customer, despite the actual wait time being held constant across the conditions.

4 - Timing Matters: Sourcing Workers in on-Demand Freight Matching Platforms

Ziqi Dong, Temple University, Philadelphia, PA, United States, Jingxuan Geng, Guangwen Kong, Qiuping Yu

We focus on the flourishing freight-matching businesses that rely on crowdsourced drivers for long-haul trucking. In particular, although the practice suggests that shippers' ordering behaviors of freight-matching services may remarkably impact crowdsourced drivers' bidding behaviors, the literature has yet to examine this issue formally. Therefore, we collect industrial data and construct a strict empirical schema for understanding the association between shippers' order timing and the freight-matching performance. Besides, by deliberately building a theoretical modeling framework and using a data-driven estimation of model parameters, we are able to simulate the freight-matching performance of adopting our empirical findings and evaluate the practical value of our study.

5 - Fairness and Efficiency Trade-offs in Multi-Class Service Systems

Hamid Arzani, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Hossein Abouee Mehrizi

In service systems, the pursuit of efficiency by prioritizing certain groups often raises concerns about fairness toward those who are being prioritized. In this paper, we study a single-server queueing system with multiple classes of customers, distinguished by their base utility, urgency level, and arrival rate. We examine two policies: the efficient (i.e., utilitarian) policy, which maximizes the system's total expected utility by prioritizing the urgent class, and the fair (i.e., egalitarian) policy, which employs a randomized priority rule to maximize the minimum expected utility of individual customers across classes. We show that the fair policy may agree or disagree with the efficient policy depending on system utilization. We then explore the efficiency-fairness trade-off by analyzing the relative efficiency loss associated with the fair policy, namely the price of fairness (POF), and the relative fairness loss associated with the efficient policy, namely the price of efficiency (POE). We illustrate that even a slight increase in service capacity can lead to significant improvements in POF and POE. We further characterize the conditions under which POF or POE is the most significant. Our results emphasize that the connection between fairness and efficiency in our model is not dichotomous; instead, it often manifests as a continuum: Substantial gain in one might be achieved with only marginal loss in the other. This implies that in situations where alternative options are limited, such as immigration or healthcare systems, a slight adjustment in service allocation can promote fairness significantly while maintaining efficiency to a large extent.

TE52

Summit - 444

Data Analytics and Technology in Operations Management

Invited Session

MSOM: Service Operations

Chair: Kenneth Moon, University of Pennsylvania, The Wharton School, Philadelphia, PA, United States

1 - Capturing Cannibalization and Complementarity Effects with Transformers

Manuel Moran-Pelaez, MIT Operations Research Center, Cambridge, MA, United States, Georgia Perakis

We present a novel approach based on Transformers for capturing cannibalization and complementarity effects within the retail sector. Traditional demand methods often assume independence between items and hence struggle to capture the intricate interplay between products. In response, we propose a framework that combines product information (textual, visual, and numerical features), and the Transformer architecture to capture product interactions. The model is validated using real-world retail data, demonstrating its ability to predict and quantify the impact of cross-item effects accurately. The findings derived from this approach enhance our understanding of cannibalization and complementarity dynamics and offer valuable insights into retail operations, such as assortment optimization, pricing, and inventory management.

2 - Structural Forecasting of Customer Arrivals to Scheduled Services

Fernando Bernstein, Duke University, Durham, NC, United States, Bora Keskin, Adam Mersereau, Morgan Wood, Serhan Ziya

Motivated by the RDU International Airport, we develop methods to forecast future arrivals to a system where arrivals are tied to scheduled events. In our scenario, passengers arrive for their scheduled flights. We consider a setting in which the historical number of passenger arrivals is known, but it is unknown for which flight each passenger arrived. Assuming each passenger arrives early for their respective flight according to some distribution, we develop methods capable of learning the arrival structure of individual passengers to their respective flights using both a parametric model and a non-parametric approach of passenger arrival earliness. Earliness distribution estimates, along with future flight schedules, are used to forecast the total number of future passenger arrivals. We explore the benefit of allowing the earliness distribution estimate to vary depending on time of day, day of week, flight-type, and holiday occurrence. We compare the performance of these passenger arrival estimates to estimates derived from machine learning methods using real-world data.

3 - High-Flying Analytics: Harnessing Wearable Sensors and AI to Safeguard Military Aviators

Xufei Liu, Wharton, University of Pennsylvania, Philadelphia, PA, United States, Gad Allon, Kenneth Moon

For military aviators, functioning without fail within a high-stress work environment is a necessity. In order to complete critical missions, they make split-second decisions while piloting 44,000 lbs of hurtling machinery accelerating at up to 9G's in high-altitude conditions. For less experienced pilots, managing the associated physical and mental fatigue is an integral and challenging component of executing flights safely and well.

Yet, currently, few measures are in place to monitor and track fatigue for pilots. The US Air Force (USAF) primarily forecasts future states of fatigue using a general-purpose biomathematical model created decades ago - we aim to bring modern data science and analytics to the problem of live fatigue management for military pilots. We collaborate with a pioneering company that innovatively outfits aviators at 22 USAF bases with wearable sensors that record both the physical stresses of sorties (flights) and the biophysical states and reactions of pilots in real-time. Then, we develop analytical methods to make timely and accurate use of such novel wearable sensor data to benefit and safeguard pilots.

Our research focuses on maximizing high-stakes, learning-dependent performance in the field of operations management. Whereas the use of sensors and testing to manage performance is far from new to the field of operations, its use in expediting expert human experience and performance in complex settings, with the aid of artificial intelligence, is still nascent. We create a data-driven solution to identify critical intervention points based on highly complex data on dynamic human performance.

4 - Designing Recommendation Algorithms to Improve Equity in Access on VolunteerMatch

Akshaya Suresh, Yale University, New Haven, CT, United States, Vahideh Manshadi, Scott Rodilitz, Daniela Saban

In collaboration with VolunteerMatch (VM)---the world's largest online platform for connecting volunteers with volunteering opportunities---we designed and implemented a new display ranking algorithm. VM's original ranking algorithm intended to maximize efficiency (i.e., the total number of connections), but as a consequence it repeatedly displayed the same few opportunities atop its ranking, effectively limiting access to volunteers for the other opportunities. To incorporate VM's desire for equity (defined as the weekly number of opportunities with at least one connection) along with efficiency, we propose a modeling framework for online display ranking in settings where it is important to manage the trade-off between the total number of connections and the equitable allocation of these connections. Taking an adversarial approach in evaluating the performance of online algorithms, we show that a class of algorithms that applies a *penalty* to opportunities after each connection provides a strong (and, in certain regimes, optimal) performance guarantee. Based on our theoretical development, we propose *SmartSort*, a simple online display ranking algorithm with a penalty term that we calibrated using VM's data and simulation. We implemented SmartSort in experiments covering Dallas-Fort Worth and all of Southern California. Using a difference-in-differences analysis, we find that the implementation of SmartSort led to an estimated 8% increase in equity (consistent across both experiments) without any significant decrease in efficiency. If SmartSort has a similar distributional effect on a national scale, an additional 30,000 connections every year will go to opportunities that would have otherwise lacked access to volunteers.

TE53

Summit - 445

Innovative Approaches to Sustainable Transportation and Workforce Diversity

Invited Session

MSOM: Sustainable Operations

Chair: Jingwei Zhang, Cornell University, Ithaca, United States

Co-Chair: Nagarajan Sethuraman, University of Kansas, Lawrence, KS, United States

1 - The Role of Driver Behavior and Interpretability in The Vehicle-to-Grid Market

Leann Thayaparan, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgia Perakis

As renewable energy production increases, energy storage becomes a significant challenge. Electric Vehicles (EVs) could act as distributed storage. However policies for EVs to charge from and discharge to the electric grid must account for complex driver behavior. In this work, we introduce a driver-focused dynamic optimization model that incorporates the EV driving behavior. The optimization incorporates multiple aspects of driver concerns including earning a dollar profit, minimizing their carbon footprint and reducing range anxiety. A concern is this setting is that drivers will not participate in discharging markets due to distrust of blackbox charging/discharging policies. We characterize charging/discharging policies that are interpretable to drivers and analytically guarantee that these policies are optimal under certain parameter regimes of the optimization. The parameter regimes translate to intuitive weightings between the multiple objectives of the optimization. We also establish that the dynamic program can be solved in closed form, addressing tractability concerns. We collaborate with an American EV manufacturer to size the potential of the discharging market. Over a twelve day test set, we find that the average driver earns $95 from engaging in the vehicle to grid market and has a carbon benefit of 0.81 tons of CO_2. This is a carbon benefit equivalent to 2,076 miles driven by the average gas powered car, 91.1 gallons of gasoline consumed, 0.102 of a home’s annual energy usage or 13.4 tree samplings grown for 10 years.

2 - Income-Based Public Transit Design: Optimizing Service and Pricing for Equitable Access

Jingwei Zhang, Cornell University, Ithaca, NY, United States, Owen Wu

With the Infrastructure Investment and Jobs Act, many public transit agencies are improving the transit system by either increasing the service level or lowering the ticket price with earmarked budget. It's unclear yet which lever is more efficient in boosting social welfare. This paper studies the transit design problem by considering different income groups' sensitivity in travel time and price. Our findings offer cost-effective prescriptions for improving urban mobility.

3 - Impact of Women in the Invention Team on Product Realization in the Pharmaceutical Industry

Nagarajan Sethuraman, University of Kansas, Lawrence, KS, United States, Deepak Jena, Rachna Shah

Does having additional women participate in research and product development teams result in more successful products downstream? We examine this question in the context of the pharmaceutical industry which has suffered from historically low women participation in the patent invention teams.

4 - Going the Distance: the Impact of Commute on Gender Diversity in Public Service

Vanitha Virudachalam, Gies College of Business, UIUC, Champaign, IL, United States, Dawson Kaaua

We analyze the extent to which longer commute distances deter female political participation and whether this effect varies according to the degree of flexibility and commensurate compensation provided by the position. Furthermore, we investigate policies that can lower the barrier to entry for women in politics. Leveraging differences in distance to the state capitol among state house districts, we show that state house districts located further from the state capitol have a lower percentage of female candidates, up to 3 to 4% with the effect plateauing at 120 miles. This effect is driven by states that provide neither a high degree of flexibility (i.e., part-time work) nor a high level of compensation (i.e., full-time work). In such states, the decrease in female participation is up to 4 to 5% with the effect also leveling off at around 120 miles. Based on two conjoint survey experiments, we find that paid parental leave might motivate women at the beginning of their political careers with longer commutes to run for office while a remote work/proxy voting policy could help to sustain the political careers of these types of women.

TE54

Summit - 446

Sustainability through Supply Chain Innovations

Invited Session

MSOM: Sustainable Operations

Chair: Luyi Gui, The Paul Merage School of Business, UC Irvine, Irvine, CA, United States

Co-Chair: Xi Lin, UC Irvine, Paul Merage School of Business, Irvine, CA, United States

1 - Sustainable Sourcing of Agricultural Products: Fixed vs. Flexible Premiums

Vishal Agrawal, Georgetown University, Washington, DC, United States, Can Zhang

Sustainability certifications have gained increasing popularity for agricultural product sourcing. To help smallholder farmers achieve a living income, these certifications set a premium that firms must pay to farmers for certified crops in addition to the market price. A popular premium approach adopted by several major certifications is called fixed premium, under which farmers receive a fixed level of premium independent of the market price. This approach has been criticized for not protecting farmers from low market prices. An alternative approach called flexible premium, under which farmers receive a higher level of premium when the market price is lower, has been promoted by farmer advocacy groups and implemented recently by some certifications. We analyze and compare the effectiveness of these two premium approaches. Our analysis reveals several insights. First, contrary to expectation, we show that flexible premium can lead to a lower expected farmer income than fixed premium. Furthermore, even when flexible premium improves farmer income, it can lead to a lower expected sustainable sourcing quantity and firm profit. Second, we nevertheless identify that flexible premium can lead to a ``win-win-win" outcome for farmer income, sustainable sourcing quantity, and firm profit, but only in ample-supply, demand-constrained settings. Finally, motivated by the trend of firms creating their own sustainability labels, we show that such self-labeling can lead to a higher sustainable sourcing quantity but still lower farmer income as compared to under NGO labeling.

2 - Labor, Skills, and Product Recovery

Cerag Pince, Quinlan School of Business, Loyola University Chicago, Chicago, IL, United States, Atalay Atasu

Motivated by a third-party remanufacturer of consumer products, we study the workforce capacity investment problem in product recovery operations. We analyze how a remanufacturing firm should determine its workforce's quantity and type when the returned product volume and complexity are random. We characterize the optimal solutions for different workforce strategies and identify the conditions when one strategy dominates the other.

3 - The Value of Advice: Evidence from Thousands of Smallholder Farms in The Philippines

Canberk Ucel, Bilkent University, Ankara, Turkey, Diwas KC

Increasing the productivity of Philippine coconut farms that are well below world standards could improve the livelihoods of 3.4 million farming families, most suffering poverty. Government and supporting organizations have long promoted Good Agricultural Practices (GAPs), which decades of public research suggests would double farm productivity with little capital investment, but have failed to achieve widespread adoption and productivity gains. We study the role of change agents in facilitating GAP adoption and effective implementation using proprietary data on the productivity, granular farming practices, and characteristics of 1,998 smallholders. Our quantitative analysis leverages the pseudo-exogenous variation in agricultural extension office locations to find that having an extension office within 8 kilometers is associated with greater awareness of central recommendations for 7 of 8 GAPs, increased adoption rates for the three most effective GAPs, and 36% higher productivity, on average, on otherwise comparable farms. Our post-hoc analysis further finds suggestive evidence that physical interactions enable agents to support complex practice adoption and implementation decisions. Moreover, we find significant heterogeneity in the effects of agent access and offer facility reallocation and visit schedules to improve service coverage and effectiveness with existing capacity. Our results suggest that supporting organizations should integrate change agent support or otherwise focus on integrating customized advice, farmer feedback, and assistance with finer details of implementation. Evidence-based provision of advisory services, extended beyond the Philippine context, could potentially benefit two billion people worldwide dependent on smallholder farms, and benefit small heterogenous firms that dominate vital functions in other industries.

TE55

Summit - 447

Applications of Operations Research and Analytics in Elections

Invited Session

Public Sector OR

Chair: Charles Thraves, University of Chile, Santiago, Chile

1 - Monotone Randomized Apportionment

Jose Correa, Universidad de Chile, Santiago, Chile, Paul Goelz, Ulrike Schmidt-Kraepelin, Jamie Tucker-Foltz, Victor Verdugo

Apportionment is the act of distributing the seats of a legislature among political parties (or states) in proportion to their vote shares (or populations). A famous impossibility by Balinski and Young (2001) shows that no apportionment method can be proportional up to one seat (quota) while also responding monotonically to changes in the votes (population monotonicity). Grimmett (2004) proposed to overcome this impossibility by randomizing the apportionment, which can achieve quota as well as perfect proportionality and monotonicity — at least in terms of the expected numberof seats awarded to each party. Still, the correlations between the seats awarded to different parties may exhibit bizarre non-monotonicities. When parties or voters care about joint events, such as whether a coalition of parties reaches a majority, these non-monotonicities can cause paradoxes, including incentives for strategic voting. In this paper, we propose monotonicity axioms ruling out these paradoxes, and study which of them can be satisfied jointly with Grimmett's axioms. Essentially, we require that, if a set of parties all receive more votes, the probability of those parties jointly receiving more seats should increase. Our work draws on a rich literature on unequal probability sampling in statistics (studied as dependent randomized rounding in computer science). Our main result shows that a sampling scheme due to Sampford (1967) satisfies Grimmett's axioms and a notion of higher-order correlation monotonicity.

2 - Maximizing Proportionality in Districting Plans

Brendan Ruskey, Lehigh Unversity Department of Industrial Engineering, Bethlehem, PA, United States, Larry Snyder

We present optimization models for generating fair redistricting plans. Every ten years in the United States, lawmakers sort the citizens of each state into groups of roughly equal population in what is called a "redistricting plan”. In the context of fairness between political parties, we define proportionality as the difference between a party's statewide share of individual votes and the share of districts in which the party has a majority of voters. In practice, disproportionate plans are seen as unfair. Therefore, we present multiple mixed-integer programming (MIP) models for the problem of maximizing proportionality within a redistricting plan. Our model differs from existing approaches, in that we present the first optimization model for maximizing proportionality assuming fractional seat shares. We also propose and implement a linear programming (LP) relaxation heuristic for the problem. We share computational results from applying our heuristic to actual US state redistricting problem instances, and compare these generated plans to existing US redistricting plans.

3 - Optimal Strategies in Ranked-Choice Voting

Sanyukta Deshpande, University of Illinois at Urbana Champaign, Champaign, IL, United States, Nikhil Garg, Sheldon Jacobson

Ranked Choice Voting and Single Transferable Voting are widely used. An open question is how candidates should strategize, forming coalitions or lending support to other candidates. Calculating optimal strategies is exponentially complex because of intricate per-round vote transfers, with minor changes in the voter rankings potentially leading to bigger ripple effects -- for example, lending support to a losing candidate can prevent their votes from transferring to a more competitive opponent. We study optimal strategies -- *persuading* voters to change their ballots or *adding* new voters -- both theoretically and algorithmically. Theoretically, we study the types of strategies that are effective under both perfect and imperfect polling information. Algorithmically, we develop efficient algorithms to reduce the election instance while maintaining optimization optimality, practically circumventing the computational complexity barrier. Finally, we apply our algorithmic approach to ranked-choice polling data on the US 2024 Republican Primary, finding, for example, that several candidates would have been optimally served by boosting another candidate instead of themselves.

4 - The *FastMap* Redistricting Algorithm: An AI Milestone for a Better American Democracy

Matthew Petering, UW-Milwaukee, Milwaukee, WI, United States

On January 12, 2024 seven proposals for Wisconsin's legislative districts were submitted to the Wisconsin Supreme Court in the case *Clarke v. Wisconsin Elections Commission*. One was generated by a computer algorithm, and six were created by expert human mapmakers. In this presentation, we examine a major development from the case: the map proposal generated by the *FastMap* algorithm significantly outperformed the other proposals. Relevant background information, an algorithm summary, and map visualizations/comparisons are provided. The case likely marks the moment when computer algorithms surpassed humans in overall mapmaking ability. Going forward, algorithmic mapmaking will enhance American democracy by ensuring that congressional, state legislative, and local election districts better reflect constitutional requirements, the will of the voters, and other criteria specified by courts and legal teams.

TE56

Summit - 448

Advanced Optimization and Modeling Techniques in Operations Research

Contributed Session

Chair: Rick Willemsen, Erasmus University Rotterdam, Rotterdam, Netherlands

1 - Random Preference Models: Estimation and Identification

Moha Ghaderi, Pompeu Fabra University, Barcelona, Spain, Milosz Kadzinski, Kamel Jedidi

Random Preference Models (RPMs) provide analytical tools to describe and predict choices directly based on preference orderings. The linear relationship between RPMs and aggregated demand, their intuitive appeal in explaining choices without relying on utility mediums, and their flexibility to capture various types of context-dependency without needing random-utility-like specification assumptions make RPMs powerful tools. On the downside, estimating RPMs requires projecting choices onto the discrete space of preference ordering with a dimensionality that grows exponentially to the number of choice options. This causes computational hurdles in estimating RPMs in a reasonable time and identification issues in pinning down the RPM parameters. We provide a unimodal representation of RPMs and explore the properties of this subclass of RPMs. We proceed by developing efficient estimation procedures that iteratively construct a restricted support for the unimodal-RPM. Moreover, we explore identification of unimodal-RPM using the standard (menu, choice) input data and provide insights into their computational and predictive performances using a simulation analysis.

2 - A fast metaheuristic for the alpha-neighbor p-center problem

Roger Rios, Universidad Autonoma de Nuevo Leon, San Nicolas de los Garza, Mexico, Ernesto Ortiz, Diana Huerta

The alpha-neighbor p-center problem (ANPCP) is a location problem in which the goal is to select p centers such that the maximum distance of a user to its alpha-th closest open facility is minimized. This problem arises when the aim is to have a minimum guaranteed response time between a demand point and its center by providing backup centers in case one of them fails to respond to an emergency. We propose a GRASP for this NP-hard problem. The construction phase evaluates the objective function of the p-dispersion problem to provide a feasible solution. We enhanced the time complexity of this heuristic by using tailored-made data structures, called Greedy Dispersion , and adapted it to use a value-based restricted candidate list within GRASP. The constructed solution is then improved, in the next phase, by a local search whose move is a vertex substitution or interchange. We adapted the concept of fast interchange specifically for the ANPCP and called it Alpha Fast Vertex Substitution. This local search is significantly faster by exploiting the structure of the objective function, allowing to reuse several expensive computations, and led to outstanding computational speed-up times for all the test cases. Empirical evidence over a wide set of instances shows the effectiveness of the proposed procedure.

3 - Variable aggregation in a Benders decomposition for the p-median problem

Rick Willemsen, Erasmus University Rotterdam, Rotterdam, Netherlands, Daniel Aloise, Raf Jans

The p-median problem is a classical discrete location problem and is equivalent to the well-known k-medoids problem in the unsupervised clustering literature. The aim is to select p centers while minimizing the sum of distances from each customer to its nearest center. Recent advancements in solving the p-median and related problems have successfully leveraged Benders decomposition methods. In order to reduce the number of variables and possibly the number of Benders cuts in these models, it is possible to aggregate distance variables corresponding to customers. We propose to partially aggregate the distance variables based on an initial solution: aggregation occurs only when the corresponding customers are assigned to the same center in the initial solution. In addition, we propose a set of tailored valid inequalities for these aggregated variables. Initial experiments indicate that our model, post-initialization, provides a stronger lower bound, thereby accelerating the resolution of the root node. Furthermore, this approach seems to positively impact the branching procedure, leading to an overall faster Benders decomposition method.

4 - Workflow Scheduling in Cloud Environments using Graph Neural Networks

Vivek Anand Rajopanth, University of Illinois Urbana-Champaign, Urbana, IL, United States, Rakesh Nagi, Asser Tantawi, Chen Wang

Workflow Scheduling is a combinatorial optimization problem that focuses on efficiently allocating user-submitted jobs to cloud processing clusters. Traditional methods predominantly employ heuristics that are swift to deploy but often deviate from achieving optimal outcomes. In contrast, Mixed Integer Programming (MIP) provides a rigorous model of the system, capable of delivering optimal solutions for any problem instance, but at the expense of more computational time. A key observation in cloud computing systems is the similarity in job types received within specific time frames. Our research capitalizes on this similarity by using a Graph Neural Network (GNN) to predict a part of the schedule (heatmap) based on historical data of successful schedules. We then use these predictions to aid a heuristic to build a feasible schedule with minimal deviations from optimality. By integrating the precision of MIP with the predictive power of the GNN, our approach significantly enhances workflow management efficiency in cloud environments.

TE57

Summit - Terrace Suite 1

System Medicine and AI: Transforming Healthcare Landscapes

Invited Session

Health Applications Society

Chair: Dimitris Bertsimas, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Yu Ma, MIT, Cambridge, MA, United States

1 - M3h: Multimodal Multitask Machine Learning for Healthcare

Yu Ma, MIT, Cambridge, MA, United States, Dimitris Bertsimas

Recent breakthroughs in AI are poised to fundamentally enhance our existing study and understanding of healthcare. Specifically, the development of an integrated many-to-many framework that leverages multiple data modality inputs for the analytical modeling of multiple medical tasks, is critical for a unified understanding of modern medicine. In this work, we introduce M3H, an explainable Multimodal Multitask Machine Learning for Healthcare framework that consolidates learning from diverse multimodal inputs across a broad spectrum of medical task categories and machine learning problem classes. We evaluate the M3H framework by validating models trained from four modalities (tabular, time-series, language, and vision) on 41 medical tasks across 4 machine learning problem classes. Our results demonstrate that M3H consistently produces multitask models that outperform canonical single-task models (by 1.1- 37.2%) across 37 disease diagnoses from 16 medical departments, three hospital operation forecast, and one patient phenotyping tasks: spanning machine learning problem classes of supervised binary classification, multiclass classification, regression, and clustering. Additionally, the framework introduces a novel attention mechanism to balance self-exploitation (focus on learning source task), and cross-exploration (encourage learning from other tasks). Furthermore, M3H provides explainability insights on how joint learning of additional tasks impacts the learning of source task using a proposed TIM score, shedding light into the dynamics of task interdependencies. Its adaptable architecture facilitates the easy customization and integration of new tasks and data modalities, establishing it as a robust, scalable candidate solution for advancing AI-driven healthcare systems.

2 - Precision Medicine in Cardiology: Enhancing Valve Prescription for Improved Outcomes

Vasiliki Stoumpou, Massachusetts Instiute of Technology, Cambridge, MA, United States, Phevos Paschalidis, Lisa Everest, Yu Ma

As a minimally invasive procedure, Transcatheter Aortic Valve Replacement (TAVR) has played an increasingly vital role to help fight the prevalent aortic stenosis condition. However, the choice of valve type has remained a much-debated question without clear guidance from either the literature or clinicians. This study leverages artificial intelligence-based techniques to tackle this debate and minimize postoperative complications, specifically permanent pacemaker implantation (PPI). We propose an interpretable, personalizable, machine-learning-based Optimal Policy Tree (OPT) approach to optimally prescribe the preferred transcatheter heart valve (THV), the Medtronic Evolut Pro Plus or Edwards Sapien 3, to minimize postoperative risks. We train our OPT model on a retrospective dataset of patients who received a first-time Edwards Sapien or Medtronic Evolut Pro TAVR procedure between January 2011 and September 2022 at a large hospital system in the US. We perform external validation of our model on a retrospective patient dataset with the same criteria, between January 2011 and December 2021 at Hygeia Hospital, Greece. On historical data from a large US hospital program, we predict a reduction of observed PPI by as much as 26%. On an external validation dataset, we predict a 16% improvement, confirming our model's generalizability on unseen data.

3 - Efficient Infection Detection: Predictive Analytics in Hospital Operations

Phevos Paschalidis, Harvard University, Cambridge, MA, United States, Dimitris Bertsimas, Carol Gao, Konstantina Rasvani

Inappropriate and excessive diagnostic testing for urinary tract infections (UTI) and bloodstream infections (BSI) presents a significant cost burden to hospitals and can result in false positives that mislead physicians and encourage overuse of antibiotics, potentially leading to adverse drug events. Optimizing the process of ordering, performing, and reporting tests such as urinalysis and blood cultures is therefore critical to ensuring high quality care upon patient admission. In collaboration with Hartford HealthCare — the largest hospital system in Connecticut — we develop two clinical support tools that estimate the risk of UTI and BSI for admitted patients. Responding to recent advancements in large-scale multimodal machine learning, we leverage both tabular and language data using the Holistic AI in Medicine (HAIM) framework. Our models achieve high AUC of 0.85 and 0.75 for UTI and BSI prediction corresponding to a 25% reduction in ordered urinalyses and 10% decrease in ordered blood cultures. Given the more than 400,000 urinalyses and 50,000 blood cultures collected annually at Hartford HealthCare, these models promise to save hundreds of thousands of dollars while accurately identifying more than 98% of infected patients.

4 - Brainsage: Learning Meaningful Representations from 3D Brain Scans For Neurological Disease Prediction

Matthew Peroni, MIT, Cambridge, MA, United States

Early detection of neurological disease is a challenging and crucial task in modern healthcare. Given the increasing availability of MRIs, especially T1-weighted (T1w) scans, there has been an effort to use brain scans for early detection and diagnosis of neurological disease. In this work, we develop deep learning models for neurological disorder classification and brain age prediction using T1w MRI scans, leveraging a large data set of scans from a major hospital network in the United States. We develop a 3D vision transformer (Vit) encoder, pre-trained as a masked autoencoder, to project the brain scans into a latent feature space that can be used for various downstream supervised prediction tasks and unsupervised tasks. We demonstrate the potential of using masked autoencoders as a pre-training technique for learning representations of the human brain, leveraging their contextual nature. The downstream models trained on the masked autoencoder representations are extremely light-weight, providing an improvement in parameter and compute efficiency compared to creating task-specific deep-learning models.

TE58

Summit - Terrace Suite 2

Data-Driven Healthcare Operations Management

Invited Session

Health Applications Society

Chair: Guihua Wang, The University of Texas at Dallas, Richardson, TX, United States

Co-Chair: Minmin Zhang, University of Texas at Dallas, Richardson, TX, United States

1 - Improving Access for Underserved Communities: Network Implications of Opening Rural Healthcare Clinics

Masoud Kamalahmadi, University of Miami, Coral Gables, FL, United States, Kurt Bretthauer, Jonathan Helm, Rodney Parker

We study the impact of new rural clinics established by the Department of Veterans Affairs on rural veterans' healthcare access and service utilization patterns. Using a difference-in-difference analysis, we find that the new clinics attracted a significant number of new patients and also increased healthcare usage among current patients, especially in rural areas. We also investigate the network effects of the new clinics on patient demand in other facilities. We find that while the new clinics partially substituted for certain services in a larger medical center, they also generated additional referrals for specialist services not available in them. Conversely, we find a decline in hospitalizations, indicating that the expansion of outpatient services may facilitate earlier diagnoses and interventions, ultimately improving health outcomes.

2 - When Less is More: An Analytical Analysis and Empirical Evidence of Control Limit Strategies in a Telemedicine Platform

Guangwen Kong, Temple University, Philadelphia, PA, United States, Jingxuan Geng, Marco Qin

We consider a telemedicine platform that allows patients to seek multiple medical diagnoses from doctors online. We find that using a commission-based pricing mechanism alone may yield a downward distortion on price to prevent the over-participation of doctors compared to a centralized benchmark. By imposing a control limit on the number of diagnoses received per inquiry, the platform can charge a higher price while maintaining the appropriate number of responses from doctors. When patients are sensitive to delays in receiving diagnoses, interestingly, a platform may benefit from a patient's increased delay sensitivity because it plays a similar role as the control limit by discouraging later arriving doctors from participating. As a result, the profit improvement from imposing a control limit mechanism may decrease with delay sensitivity. When doctors are heterogeneous in their service quality, the undesirable outcome of low-quality doctors driving out all high-quality doctors may occur. A control limit could increase high-quality doctors' participation by increasing their chances of being rewarded by patients. We empirically test the predictions using data from a large telemedicine platform and find that it supports the results from the model analysis.

3 - A Multi-Treatment Forest Approach for Analyzing the Heterogeneous Effects of Team Familiarity

Minmin Zhang, University of Texas at Dallas, Richardson, TX, United States, Guihua Wang, Wallace Hopp, Michael Mathis

This study investigates the impact of team familiarity on surgery duration, exploring whether this effect varies among patients. Introducing and applying the "MT forest" approach to orthopedic surgeries, we show that increased team familiarity, particularly between anesthesiologists and nurses, and surgeons and anesthesiologists, significantly reduces surgery duration. Moreover, the effect of team familiarity varies across patients with different characteristics. By developing an optimization model, the study suggests that leveraging these variable effects enhances hospital operational efficiency by better aligning surgical teams with patients.

4 - The Impact of Inspection Backlogs on Nursing Home Quality: State Agencies&Rsquo; Lax Oversight and Staffing Shortage

Wilson Lin, Santa Clara University, Santa Clara, CA, United States, Susan Lu, Lauren Lu

This study examines the impact of inspection delays on nursing home operations. Using nursing home data from 2013 to 2023 from the Centers for Medicare & Medicaid Services, we find that inspection backlogs are associated with increased nursing home deficiency citations. These effects become stronger as these inspection delays beyond the statutory 15 months between standard health inspections. Furthermore, we find that the effects of these delays are heterogeneous. These findings facilitate us to design a new inspection mechanism under insufficient staffing supply.

TE59

Summit - Ballroom 1

AI and Environmental Sustainability Innovations

Contributed Session

Chair: Carlos Olivos, Universidad Católica del Norte, Avenida Angamos 0610, Antofagasta, Chile

1 - Harnessing AI for Holistic Emissions Monitoring and Management in Nature-Based Climate Solutions

Seyed Shahabeddin Mousavi, Stanford University, Stanford, CA, United States, Olamide Oladeji

Amidst the recent rapid advances in the domain of artificial intelligence (AI) and the exploration of possibilities for how AI can change industries, AI has been proposed as an important tool to support an array of efforts related to nature-based climate solutions, including but not limited to the forecasting and detection of wildfires that affect forests and vegetation-based offsets. While this application and others provide important demonstrative value for the extent to which AI can impact climate change mitigation efforts, such efforts have typically been undertaken in disjointed silos, lacking sufficient awareness of the integrative nature of real-world climate policy-making. We propose a novel overarching framework for AI-powered integrated and comprehensive decision support for the various aspects of nature-based climate decision-making. With a focus on vegetation-based solutions such as forests, we demonstrate how different AI-aided decision support models such as AI-powered wildfire detection, vegetation carbon stock assessment, reversal risk mitigation, and disaster response planning can be integrated into a comprehensive framework. Rather than viewing these models as disparate elements working independently, we posit that the exchange of data and analytical results across elements of the framework as they function in tandem and careful mitigation of uncertainty propagation provide tremendous value and significant improvements relative to the status quo for real-world climate policy-making. This comprehensive framework is meant to be utilized by decision-makers to provide comprehensive computational decision support for nature-based climate solutions, enabling steadfast advancement in the domain through scientific backing for climate targets and proposed policies.

2 - Equipping Engineering and Science Students for the Modern Age of Artificial Intelligence

Stephany Coffman-Wolph, Ohio Northern University, Ada, OH, United States

We are currently in an era of unprecedented advancement of artificial intelligence (AI) and the integration of AI into various facets of everyday life. A comprehensive understanding of AI fundamentals is essential to equip the next generation of engineers and scientists with the necessary skills to navigate this ever-changing field. This presentation will cover the curriculum for a course that provides students with a holistic overview of AI and AI fundamentals. By empowering students with AI literacy, we aim to cultivate a generation of innovators equipped to leverage AI technologies responsibly and ethically. Ultimately, we aim to inspire and prepare tomorrow's leaders to harness AI's power for society's betterment.

3 - A bi-level programming approach to climate policy design in liberalized power systems under risk

Emil Dimanchev, Norwegian University of Science and Technology, Trondheim, Norway, Steven Gabriel, Stein-Erik Fleten, Filippo Pecci, Magnus Korpas

Climate policy makers must incentivize risk-averse firms to deploy large amounts of capital toward clean energy. In liberalized electricity markets, firms can face considerable risks due to the incompleteness of long-term markets, which is also known as the missing market problem. We consider how popular climate policies should be optimally designed in such markets. Our focus is specifically on investment tax credits, such as those in the U.S. Inflation Reduction Act, and carbon taxes. We introduce a new analytical framework for modeling a government's optimal policy decisions in anticipation of electricity market behavior under risk. Our model uses bi-level programming to endogenize subsidy choice, unlike traditional models which treat such policies as exogenous. A simple case study demonstrates our approach for an abstract power system with a traditional gas technology, variable renewables, storage, and a clean firm technology, under demand and gas price uncertainty. In illustrative experiments, we observe that optimal investment tax credits and the optimal carbon tax are both higher when long-term markets are missing than when markets are complete. In some of our cases, investment tax credits are more cost-effective than a carbon tax. This occurs because, by increasing investment in renewables, subsidies can mitigate unhedged societal risk from gas price uncertainty.

4 - Dynamic ESG Scoring Modelling in Asset Management

ziru tao, University of Manchester, manchester, United Kingdom, paul baguley, clara cheung, obuks ejohwomu

In the pursuit of a better world and the United Nations Sustainable Development goals, smarter decisions are now critical in the realm of engineering project management. The integration of Environmental, Social, and Governance (ESG) factors into project cost estimation is now critical and timely, and Artificial Intelligence (AI) presents a transformative tool for this task. This research **Aims** to develop an AI-assisted methodology to estimate ESG-related project costs for facilities management of fleets of buildings, disrupting traditional cost management approaches and enhancing decision-making for a more sustainable future.

The **Research Methodology** includes a state-of-art review of 108 journal papers used to synthesise an initial conceptual framework. It includes Estimating Requirements, Data Integration, ESG Costs and Cost Drivers, AI Architecture, and the critical aspect of AI Explainability. The objective of the framework is to guide smarter decisions by providing a clear, accountable, and data-driven approach to ESG cost estimation, thereby contributing to a better world through responsible project management.

The **Findings** are an analysis of a global facilities management company's estimating requirements, and a dynamic ESG scoring model using Recurrent Neural Networks (RNN) to evaluate the ESG performance of fleets of buildings during the use stage of the full life cycle.

The prediction results are favourable in that a chosen ESG metric has been reasonably predicted for the buildings in the data set. However, **Limitations** exist in the scope of the data set analysed, the subjectivity in the ESG metric developed, and the lack of synthesis with private cost data.

5 - Integrating Battery Energy Storage Systems for Renewable Energy Deployment

Carlos Olivos, Universidad Católica del Norte, Antofagasta, Chile, Javier Flores

The Unit Commitment Problem (UCP) is a classical problem in the Operations Research and Power Systems field. During recent years, there has been a significant interest in phasing out fossil fuel-based energy and transitioning towards renewable sources. However, the intermittency of renewables yields to fluctuations, resulting in infeasible pre-dispatch schedules. Battery Energy Storage Systems (BESS) is a technology that could ease the energy transition by storing the surplus of renewable energy and discharging it when energy from renewable sources is not available. In this study, we integrate BESS into the UCP to address renewable energy variability. We evaluate various BESS materials and their impact on the feasibility and cost of energy schedules. Results, conclusions, and algorithms will be presented.

TE60

Summit - Ballroom 2

Resilient and Equitable Power Grid Management

Contributed Session

Chair: Juliette Ugirumurera, National Renewable Energy Laboratory, Golden, CO, United States

1 - Envisioning the Future of Power System Operations: An Energyshed Framework

Dakota Hamilton, University of Vermont, Burlington, VT, United States, Samuel Chevalier, Amritanshu Pandey, Mads Almassalkhi

In an effort to enable green energy transition and increase energy resilience, local and regional governments around the world have begun to incentivize the creation of local energy communities through public policy decisions. In this work, we explore the concept of energysheds, which are contiguous energy communities in which a minimum percentage of local energy demand is met by local energy generation over a given window of time, and investigate the impact of energysheds on future power systems operations. We introduce a mathematical framework for studying such energyshed operations and formulate energyshed policy design as an optimization problem which enables communities to achieve their energy transition goals by managing operational flexibility in an equitable and cost-effective manner. The non-linearity of such local energyshed policy requirements and the underlying power system physics results in non-convex optimization problems. Despite this, we demonstrate how the problem structure can be exploited by leveraging techniques from quasi-convex optimization and parametric optimization in order to find globally optimal solutions within arbitrary precision. Through this framework, we study how physical power system infrastructure and public policy requirements can enable or hinder the ability of local energy communities to achieve their energy goals. Finally, we demonstrate the importance of cooperation and coordination across energysheds, which are interconnected by electrical networks, as a critical step in green transition.

2 - Power grid reliability during future extreme heat and wildfire events

Hannah Burroughs, Lawrence Livermore National Laboratory, Livermore, CA, United States, Juliette Franzman, Jhi-Young Joo, Jean-Paul Watson, Andrew Mastin

Heat waves can have significant human and economic impacts including strain on the power grid. In addition, if wildfires happen concurrently with heat waves, they can further strain the power grid due to requirements to de-energize power lines near active fires or electrical infrastructure damaged within wildfire perimeters. As the probability and severity of concurrent heat waves and wildfires increases due to global climate change it is insufficient to use historical data to understand the potential impacts of these extreme events.

We develop a methodology to study the power reliability impacts of future concurrent extreme heat and wildfires. First, we identify plausible future heat wave events using temperature data from a high-resolution global climate model. The temperature data set is then used to predict increase in electrical load. In parallel, we identify future high risk wildfire areas using global climate model projections. To determine the power grid impact of plausible future wildfire events we use a bi-level optimization method. The objective is to find the set of power grid components which, if damaged by wildfire, will lead to the largest loss of load, with an understanding of how the grid operator will adjust the system in the event of outages. Finally, we use the hourly electrical load data and wildfire contingency set (electrical components disable during fire) to study the reliability impacts on the power grid (using a steady state AC power flow) from these concurrent events.

3 - Equitably Allocating Wildfire Resilience Investments for Power Grids: The Curse of Aggregation and Vulnerability Indices

Madeleine Pollack, Massachusetts Institute of Technology, Cambridge, MA, United States, Ryan Piansky, Swati Gupta, Alyssa Kody, Daniel Molzahn

Wildfires from power lines are devastating, with recent fires like the Camp Fire in California and the Smokehouse Creek Fire illustrating the need for effective, long-term solutions to fires ignited from power infrastructure. Emergency power shutoffs reduce wildfire risk from power lines; however, network, geographic, and demographic factors can lead to disproportionate power lost by vulnerable groups. Hence, power line undergrounding is an attractive, if costly, alternative. In this study, we model and evaluate Biden-Harris Justice40 policies for equitable power shutoff and undergrounding decisions on a high-fidelity synthetic transmission network in Texas using a Mixed Integer Program. We find that vulnerability indices suffer from “the curse of aggregation,” preventing targeted intervention to at-risk groups. Therefore, we recommend coupling vulnerability indices with percentage-based “min max fairness” objectives to prioritize relief to these groups.

4 - New integer programming formulations for parallel power system restoration

Hatice Çalık, KU Leuven, ELECTA, Department of Electrical Engineering, Leuven, Belgium, Dirk Van Hertem

Blackouts are emergency situations with tremendous economic and societal impacts. To minimize these impacts, efficient restoration of power systems after a blackout is of extreme importance. We investigate novel methods for restoring power systems as quickly as possible after a total blackout. Particularly, we introduce new integer programming formulations for the NP-Hard generator start-up sequencing (GSS) and parallel power system restoration (PPSR) problems. A blackout refers to a complete loss of power in the grid. Only a subset of generators has the capability to restart themselves to quickly reach their maximum capacity and energize other generators and critical loads. Such generators are referred to as ‘black start’ (BS) generators. BS capacity is often not sufficient to energize the entire network directly.

Energizing one or more non-BS generators of higher capacity gradually increases the power capacity and step-by-step restores the entire network. The GSS is challenging since each non-BS generator requires an initial cranking power for a certain period. Once the cranking period is over, the power output of the generator increases at a certain rate until it reaches the maximum capacity. The PPSR combines network design decisions with the GSS to enable partitioning the network into connected components for safer restoration. Our models provide tight bounds for both problems and they outperform the state-of-the-art formulations in terms of speed on several IEEE network instances. We also discuss model extensions for tackling uncertainty due to high penetration of renewable energy resources.

5 - Enhancing Power Grid Resilience: a Stochastic Programming Approach for Infrastructure Hardening

Minhang Zhou, University of Arizona, Tucson, AZ, United States, Pavlo Krokhmal

Climate hazards such as hurricanes pose significant threats to overhead power distribution systems, demanding improved resilience under extreme weather conditions. To address this, we propose a two-stage stochastic programming model aimed at optimally hardening power grid infrastructure. In the first stage, the model determines the hardening degrees for each transmission line, while in the second stage, it seeks to minimize total load shedding and assign optimal power flows across the lines. A novel aspect of our model is its handling of decision-dependent uncertainty, specifically the survival probabilities of distribution lines, which vary based on first-stage hardening decisions. We employ sample average approximation (SAA) to manage problem size and a modified Benders decomposition algorithm to enhance computational efficiency. The validity and effectiveness of our approach are demonstrated through application to the IEEE 33-bus distribution system under typical and extreme scenarios. This study provides a strategic framework for resource allocation in power systems facing increasing risks from climate-related events.

6 - Energy Justice Constrained Power System Dispatch and Planning

Juliette Ugirumurera, National Renewable Energy Laboratory, Golden, CO, United States, Sarah Awara, Devon Sigler, Vikram Ravi, Yun Li, Sarah Inskeep

Power plants are a major contributor to air pollution, and our current grid modeling systems lack the ability to integrate air pollution and health impact considerations from the electricity industry into their optimization processes. To address this gap, we propose a novel approach that connects a power system model (Sienna) with a high-fidelity air quality model (SCICHEM) and a human health impact assessment tool (BenMAP). This integrated system evaluates the health and damage costs associated with operating power plants and can identify environmental justice impacts on different demographic groups, especially vulnerable communities.This approach will also guide better decisions in capacity expansion models to avoid disproportionately affecting certain communities in future energy systems. A case study will demonstrate the modeling capabilities and provide example analyses for regions with at-risk populations. This new system can be utilized by system operators, researchers, and other stakeholders to improve energy planning and operations that ensure equity and understand the effects of our current power system on specific communities.

TE61

Summit - Ballroom 3

OR in the Utilities Industry

Invited Session

The Practice Section of INFORMS

Chair: Kathryn Walter, Avista Corporation, Spokane, WA, United States

1 - Implementing Optimization at Avista

Kathryn Walter, Avista Corporation, Spokane, WA, United States

Over the past dozen years, Avista has gone from not using mathematical modeling to having customized optimization models embedded within the organization to help with both short term and long term decisions. This presentation will discuss how the team grew the Avista Decision Support System (ADSS) from one person’s idea to a highly used and highly impactful tool. The audience will learn what optimization for the utilities industry looks like in practice.

2 - *Optimization-Based Framework for Enhancing Photovoltaic Hosting Capacity with Battery Energy Storage under Uncertain PV Adoption*

Khash Mahani, Quanta Technology, Brooklyn, NY, United States

Determining photovoltaic (PV) hosting capacity and enhancing it while considering uncertainties surrounding PV adoption is essential for renewable promotion and robust grid planning and operations. Existing methods for determining PV hosting capacity under uncertain PV adoption have limitations in terms of computational efficiency, accurate representation of probabilistic PV adoption, and applicability for designing solutions to increase hosting capacity beyond current limits. This presentation focuses on an optimization-based probabilistic integrated planning framework for distribution grids. It assesses feeder hosting capacity while assuming uncertain locations and amounts of PV adoption and optimizes battery deployment to increase hosting capacity limits to desired targets under those probabilistic scenarios. The framework is then applied to a distribution feeder, validating the methodology and demonstrating the benefits of batteries in maximizing PV penetration. The proposed framework can provide guidance to planners in making informed decisions regarding PV deployment strategies and effective utilization of batteries for renewable integration.

3 - Representing future hydrogen markets in the National Energy Modeling System

Stephen York, U.S. Energy Information Administration, Washington, DC, United States, Kathryn Dyl

The U.S. Energy Information Administration (EIA) will discuss our efforts to develop and integrate a representation of future hydrogen markets within our National Energy Modeling System (NEMS).

The NEMS model produces multi-decadal, long-term projections of energy systems for the United States as published in the Annual Energy Outlook (AEO). NEMS produces a general equilibrium solution for U.S. energy markets on an annual basis and includes detailed representation of energy demand, transformation, and supply projections.

A new Hydrogen Market Module (HMM) will represent the domestic hydrogen market in AEO2025. Representing an integrated hydrogen market in the NEMS allows us to analyze the potential growth in hydrogen use as a clean energy source and to reflect current laws and regulations. It also allows us to evaluate the mid- to long-term impacts of these regulations, such as the Section 45V hydrogen production tax credits in the Inflation Reduction Act, and to evaluate the possible roles of hydrogen in deep decarbonization scenarios.

The HMM projects hydrogen production by technology, connecting supply to consumers. HMM demands electricity and natural gas, transforms fuels into hydrogen, and delivers it to consumers. For AEO2025, HMM focuses on representing a simplified, integrated U.S. hydrogen market reflecting key dynamics. We are simultaneously modifying existing NEMS consumption representation for ammonia plants and refineries and adding new potential sources of hydrogen demand for the industrial, transportation, and electric power sectors.

The EIA is a U.S. statistical agency that collects, analyzes, and disseminates independent and impartial energy information.

TE62

Summit - Signature Room

Role of Influencers in Social Media

Contributed Session

Chair: Yigit Ege Bayiz, 2207 Leon Street, 78705

1 - Recommendation Agents: Transformer-Based Personalized Recommendation Systems

Landon Butler, University of California, Berkeley, Berkeley, CA, United States, Yigit Efe Erginbas, Andrew Qin, Kannan Ramchandran

The recommendation systems employed in large online marketplaces such as Amazon and Alibaba are typically designed in-house, resulting in recommendations that are geared towards maximizing the platform's profits rather than prioritizing consumer welfare. Third-party services such as Google Shopping exist to aggregate products from various online retailers, but generate revenue through paid product promotions. All in all, by using recommendation services whose goals are not fully aligned with their own, a consumer may encounter difficulties in finding the product that best meets their preferences. Furthermore, it offers little incentive for the consumer to offer feedback that improves the recommendation services for themselves and others.

Building on the recent emergence of powerful AI personal assistants, we introduce recommendation agents: interactive, cross-domain recommendation algorithms seeking to maximize the welfare of a particular consumer. Our agent comprises three sub-modules: a sequential recommendation transformer, which learns the consumer's evolving preferences from historical purchase sequences; a personalized recommendation constructor, which accounts for the consumer's idiosyncrasies such as price and risk sensitivities; and an Augmented Language Model, designed to explain recommendations and integrate consumer feedback. Through testing on a large-scale Amazon dataset, we demonstrate empirically that our agents can accurately learn consumer preferences and offer strong predictive power. We conclude with discussion on the construction of markets over many recommendation agents, allowing for efficient price discovery in large online marketplaces.

2 - Who Are the Favorites of Multi-Channel Networks? Influencer Selection with Product Collaboration

Xiongfei Guo, University of Science and Technology of China, Hefei, China, People's Republic of, Yangyang XIE, Jie Wu

We consider a brand cooperating with a Multi-Channel Network (MCN) to sell product on sales events. The brand may provide standard product, or utilize MCN’s IP into its product, called collaboration product. The MCN works with multiple heterogeneous influencers differing in performance and type. We investigate the brand’s product policy, MCN’s influencer selection strategy, and influencers’ pricing decisions with standard and collaboration products. We find that product collaboration achieves better utilization of influencers’ performance and better coverage of market, which always benefits the brand and may hurts the MCN. We obtain the MCN’s optimal influencer selection strategy, which trades off between matching influencers’ type with consumers’ taste and avoiding intensified influencer competition, and provide a quick judgement for adopting the multiple-influencer strategy. We also reveal that with standard product, a higher commission rate is always preferred by the MCN, whereas with collaboration product, making a concession to the brand in the commission rate may benefit the MCN. Additionally, higher-performance influencers may hurt brands providing standard product, while always benefit brands with collaboration product. Meanwhile, higher-performance influencers may be non-favorable for an MCN with both products. Technically, we develop a framework dealing with game involved optimization problems where infinite equilibria exist.

3 - Impact of Average Rating and Number of Reviews on Business Performance on Digital Platform

Samayita Guha, Florida International University, Miami, FL, United States, Naveen Kumar, Subodha Kumar, Joydeep Srivastava

Consumers increasingly use social media contents to make their purchase decision. A significant number of consumers study online reviews about products or services before purchasing. Online review platforms, such as Yelp and TripAdvisor, provide consumers with reviews and ratings of businesses and service providers. Hence, all types of businesses maintain their presence on online review platforms to improve their performance. To make their presence felt strongly among so many other competitors on online review platforms, the businesses need to understand which review characteristic (such as volume, rating value, reviewer identity, texts, etc.) has more influence on consumer purchase decision and in what ways. In this context, it becomes important to understand the interplay between both review rating and review number. To study this interplay empirically, we are using a data set about restaurants that are present on Yelp platform. The answer to our research question will help restaurants in allocating their resources more effectively to improve the number of reviews and improve their rating, depending on whether they have high or low average ratings.

4 - A Dynamic Model of Influencers' Organic and Sponsored Content Creation

Michael Yang, University of Texas at Dallas, Dallas, TX, United States, Xingchen Xu, Yong Tan

In online video platforms, influencers draw their audience by consistently generating high-quality content. Concurrently, they engage in the production of sponsored content, which provides direct economic benefits. This approach, however, incurs potential costs as followers may respond negatively to overt commercialization, impacting the influencers' long-term earning potential. Influencers must therefore strategically balance immediate financial returns with the dynamics of follower engagement to sustain future gains. Additionally, the manner in which sponsored content is presented—including its placement and duration—significantly influences viewership metrics and follower behavior. In light of these dynamics, we have developed the first dynamic model that accounts for forward-looking and strategic content creation decisions among influencers. This model has been calibrated using data from a prominent Chinese video platform. We also propose a counterfactual policy designed to optimize content delivery efficiency.

5 - Ranking Refinement for Increasing Credibility in Social Media Recommendations

Yigit Ege Bayiz, The University of Texas at Austin, Austin, TX, United States, Arash Amini, Ufuk Topcu

Social media often includes content derived from news sources. However, these sources are not always credible, which negatively impacts the credibility of the social media content. We present an optimization-based method designed to improve the source credibility of news content on social media platforms by refining existing content feed rankings to include source credibility scores. Our proposed solution uses a dual-objective optimization approach that minimizes the Kendall tau distance to the original ordering to approximately maintain the original content order while incorporating a linear cost function tailored to elevate the credibility of sources. We showcase our approach through an experimental setup using Reddit posts, where our algorithm doubles the credibility of ranked content compared to Reddit's hot ranking algorithm. Additionally, the algorithm allows for the incorporation of arbitrary scores other than source credibility, making it applicable across various platforms and capable of refining any arbitrary ranking of items with an additional linear cost objective.

TE63

Regency - 601

AI for Social Good

Invited Session

Information Systems

Chair: Aaron Cheng, London School of Economics, London, United Kingdom

1 - Roleplaying of AI: the Effect of Authoritative and Altruistic Ai on Promoting Prosocial Behaviors

Haoze Wu, University of Rochester, Rochester, NY, United States, Junyuan Ke, Weiguang Wang

In the implementation of AI, identity plays a critical factor, with human and AI identities exhibiting varied effects. However, existing literature on AI identity typically discloses the tool’s nature as AI rather than portraying it as human. Within this framework, different AI roles can be assigned, but their effects remain unclear. This study explores how varying AI roles impact effectiveness due to their assigned identities. We conducted a field experiment at a business hotel to investigate the effect of AI systems fulfilling different roles on environmental sustainability, specifically water consumption reduction. Based on our AI’s personalized prediction on water usage for different customers, messages were generated with varying degrees of authority and altruism to reducing water usage. Based on our randomized field experiment of the four study groups, no AI, neutral AI, authoritative AI, and altruistic AI, the findings reveal that authoritative AI messaging significantly reduces water consumption among guests. In contrast, no significant reduction is observed for neutral and altruistic AI, compared with no AI. The study underscores that AI’s persuasive capabilities vary with the different perceived roles. Overall, this research contributes to the broader discourse on AI’s role in human-AI collaboration and resource management.

2 - The Impact of Artificial Intelligence (AI) Powered Micro-gym on Public Health Welfare

Siliang Tong, Nanyang Technological University, Singapore, Singapore, Mingjie Ma, Kimhuat Goh, Yixing Chen

The rise of AI empowers the prevalence of micro-gym, a compact gym center well connected through AI technology for exercise booking and movement monitor and feasible for single person private exercises. The emergence of AI-powered micro-gym not only facilitates exerciser with better understanding of their exercise progress with AI data tracking and analysis, but also complement the discrepancy between commercial gym center and public gym facilities in local neighborhood with the flexibility and accessibility. In this study, we investigate the social impact of AI micro-gym rollout on public health welfare improvement. Leveraging a granular doctor visit record database across subzones in Singapore, we gauge how the introduction of the micro-gym affects the likelihood of doctor visits of local residents. Our empirical analyses with a stagger Difference-in-Differences framework demonstrate that the introduction of micro-gym significantly reduces the likelihood of doctor visits for local residences, indicating an improvement of the public health welfare. Further analyses across different types of diseases suggest that the introduction of micro-gym in local subzones reduces the likelihood of doctor visits for chronic diseases, cancer, digestive diseases, endocrine, musculoskeletal, mental diseases, stroke, and physical injury.

3 - Safety on Autopilot: An Empirical Investigation of Autonomous Driving and Traffic Safety

Jiyong Park, University of Georgia, Athens, GA, United States, Miyeon Jung, Min-Seok Pang

While autonomous driving holds the promise to confer economic and societal benefits, there has been a persistent concern on the safety of vehicles on autopilot. This study investigates the impact of automated driving systems on traffic safety. We design a natural experiment by combining Tesla’s rollout of ‘Navigate on Autopilot’ (NoA) in the U.S., an advanced autopilot feature that enables semi-autonomous driving on the highway, with regional variations in the intensity of Tesla vehicles. Using granular-level data on traffic accidents and electric car registrations in the Washington state during the period 2011–2022, our findings show that the intensity of Tesla vehicles is significantly associated with the reduction in the number of traffic accidents after the rollout of NoA, especially for non-fatal accidents. Additionally, it has proven more effective in preventing single-car accidents involving pedestrians or collisions between Tesla and non-Tesla vehicles, further substantiating the safety advantages of this technology. Moreover, we reveal the circumstances and conditions under which vehicles on autopilot can enhance traffic safety. Specifically, the semi-autonomous driving feature of Tesla demonstrates its effectiveness in preventing non-fatal accidents, especially under circumstances with limited visibility, less familiar routes, and less uncertainties on the roads. Taken together, our findings imply that the safety implications of autonomous driving are shaped by situational factors, encompassing not only the constraints of human capabilities but also the technological capabilities that can overcome these challenges. This study sheds new light on the potential of automated systems to prevent human errors and enhance safety.

TE64

Regency - 602

Machine Learning and Game Theory in Policy and Sports

Contributed Session

Chair: William Melville, Brigham Young University, Brigham Young University, Provo, UT, 84602, United States

1 - Integrating Machine Learning with System Dynamics: A Pioneering Approach to Addressing Policy Leakage and Enhancing Environmental Governance <cib-overlay></cib-overlay>

Satwinder Singh, Indian Institute of Management Amritsar, Amritsar, India, Aarushi Jain

In the realm of Public Policy research, the mitigation of policy leakage is a critical concern. This study introduces an innovative methodology that amalgamates machine learning with System Dynamics, thereby augmenting the analytical prowess. By harnessing the insights derived from machine learning, System Dynamics is able to employ optimal feature selection, thereby streamlining the analysis of voluminous datasets. This hybrid model significantly bolsters analytical accuracy and efficiency, which in turn enhances decision-making processes across a multitude of domains. The integration of machine learning techniques enables System Dynamics to adeptly navigate the intricacies of large datasets, thereby facilitating more profound insights and informed analyses. Furthermore, the employment of optimal feature selection methodologies ensures the procurement of relevant data, thereby amplifying the capacity of System Dynamics to provide meaningful insights. This approach ultimately provides an efficacious solution for the analysis of extensive datasets, obviating the need for manual intervention and economizing resources. Consequently, this advances decision-making processes across a diverse array of domain of policy leakage research.

2 - Analysis on relationship between language models and data quality through personality type prediction

Sungmin Lim, Ewha Woman's University, Seoul, Korea, Republic of, Juheon Kwak, Dongil Kim

In social media data analysis, the Myers–Briggs Type Indicator (MBTI) is used as a label for data analysis and artificial intelligence model training. Although the format and interests of posts may vary depending on the platform, there is research showing that an individual's text input style and speaking style revealed in the context of the post is related to MBTI. In this study, we evaluate the performance of several transformer-based language models using a low-capacity, high-quality Kaggle dataset. Then, we conducted the same experiment with high-volume, low-quality Twitter MBTI data and Reddit MBTI data to analyze which model using the data had better performance and even the performance when cross-verified.

3 - A Game Theoretical Approach to Baseball Managerial Decisions

William Melville, Brigham Young University, Provo, UT, United States, Tristan Mott, David Grimsman, Christopher Archibald

During baseball games, managers are tasked with making important strategical decisions such as when to replace a starting pitcher with a reliever, when (and with whom) to pinch hit for a batter, and when to make a defensive substitution. An optimally-strategic manager must consider how the opposing manager might react to their decisions. For instance, a manager may hesitate to bring in a relief pitcher to face an opposing hitter if they know that their opponent is likely to replace that hitter with a pinch hitter. In this work we model these baseball matchup decisions as an extensive-form zero-sum game played between two baseball managers. The goal of each manager is to take the sequence of actions that leads their team to a win with the highest probability. As part of our game model we present a novel approach to calculating the probability of plate appearance outcomes between a batter and a pitcher. This model considers the skill of the fielders and the effect of pitch count, which helps identify when a pitcher is too fatigued to continue pitching. We conclude that a Nash equilibrium strategy in this game identifies a baseball management strategy that is robust to opposing counterstrategies.

TE65

Regency - 603

Artificial Intelligence in Banking

Invited Session

Artificial Intelligence

Chair: Huan Yu, University of Southampton, Southampton, United Kingdom

Co-Chair: Cristian Roman, Western University, London, ON, Canada

1 - The Impact of Supply Chain Networks on Correlated Default for SMEs

Sahab Zandi, Western University, London, ON, Canada, Kamesh Korangi, María Óskarsdóttir, Christophe Mues, Cristián Bravo

In the domain of financial risk assessment, some recent research has focused on understanding how credit risk spreads among Small and Medium-sized Enterprises (SMEs). SMEs often form complex networks through mutual dependencies and financial interactions, which dynamically evolve. This study utilizes multilayer network data from SMEs, provided by a prominent financial institution, to estimate credit risk. Our approach leverages graph neural networks to predict SME default, using multilayer network data where connections are based on company ownerships and financial transactions between them. We demonstrate that combining this information with traditional structured data enhances application credit scoring performance, and explicitly models contagion risk across companies. Moreover, the study explores how the directionality and intensity of these connections influence risk propagation, offering a more profound understanding of the underlying processes. Our findings, validated against multiple baseline models, reveal the significance of incorporating network data and underscore the critical role of supply chain networks in the correlated default risks faced by SMEs.

2 - Breaking Barriers: Unveiling Gender Disparities in Corporate Board Career Paths using Deep Learning

Yuhao (Jet) Zhou, Western University, London, ON, Canada, Collins Ntim, María Óskarsdóttir, Matt Davison, Cristian Bravo

This study delves into the interplay between gender, professional networking, career-path trajectory, and board director appointments in Canadian publicly traded companies. At the core of our inquiry lies a uniquely constructed dataset, meticulously derived from BoardEx. This dataset (consisting of over 700 Canadian firms covering more than 19,000 senior managers and board members between 2000-2022), presents detailed network connections of both senior managers and board members across five key dimensions: education, current and prior employment, and current and prior social engagement. By matching senior managers of both genders based on their career trajectories and backgrounds, and applying Long Short-Term Memory (LSTM) deep learning alongside network analysis, we uncovered the distinct network influences impacting the board appointment prospects of women versus men. The findings causally demonstrate a "glass ceiling", suggesting that women necessitate more substantial networking credentials than men to secure equivalent corporate board positions.

3 - Self-explainable Nowcasting with Large Language Models

Daniel Abib, University of Western Ontario, London, ON, Canada, Cristian Bravo, Sanghyun Jung

We build a novel multimodal learning model with the capability of generating self-explanatory documentation.

Using separate blocks for forecasting economic time series and generating text, we use a Large Language Model (LLM) to fine tune the desired combined output to be a forecasting report with both the input numbers and a comment resembling an expert forecaster. The complexity of the document can then be tweaked according to necessity.

The model being built in blocks allows for using the preferred forecasting method, as well as for updating the LLM model when a superior one becomes available by simply switching the blocks.

4 - Score2Text: a Large Language Model Designed for Interpretable Credit Scoring

Huan Yu, University of Southampton, Southampton, United Kingdom

The widespread adoption of predictive modelling in finance institutions has significantly improved credit risk assessments. Traditional credit scoring models has been criticized of lack the transparency and interpretability for stakeholders to understand the rationale behind credit decisions. This study explores the innovative use of large language models (LLMs) to generate natural language explanations for credit scoring. By leveraging the advanced nature language processing capability model of LLMs, we aim to produce clear, human-readable interpretations of credit risk predictions. Our methodology integrates fine-tuned Llama with established interpretability technique of SHAP (SHapley Additive exPlanations) to elucidate the factors influencing credit predictions. Experimental results on diverse credit scoring datasets demonstrate that our approach not only significantly enhance interpretability of credit scores but also makes it possible to automate credit analysis reports. This study highlights the potential of LLMs to provide transparent and comprehensible explanations, fostering greater trust and enabling more informed decision-making in financial institutions.

5 - Finding paths to financial wellness: comparing investment decisions through time

Sanghyun Jung, Western University, London, ON, Canada, Cristian Bravo

The financial decision-making process of individuals is highly complex, influenced by a myriad of factors including demographics, income, macroeconomic conditions, and previous decisions. Traditional research often models this process by focusing on a limited number of these factors, which falls short of capturing the complexity of reality and fails to predict counterfactual outcomes of decision-making pathways accurately. With the rapid advancement of deep learning, there is now potential to leverage extensive datasets to more accurately model complex investment decision-making processes. However, deep learning models often struggle with discerning potential outcomes along specific decision paths. In response, our study introduces a novel causal deep learning model that leverages a Causal Transformer architecture, enhanced by incorporating time-fixed transformer blocks. These blocks compute attention scores for time-fixed variables that universally affect observations but not vice versa. This model is designed to estimate and interpret the causal impact of investment decisions. Utilizing data such as Know Your Client (KYC) profiles, account transactions, and macroeconomic indicators, we analyze which factors significantly influence decision-making paths. Moreover, by exploring counterfactual scenarios, we determine which decision paths could lead to desired financial wellness outcomes for individuals. This approach enhances our understanding of complex decision dynamics and aids in the practical application of achieving targeted financial objectives.

TE66

Regency - 604

Adoption and Applications of AI Agents Featuring ML Algorithms

Invited Session

Artificial Intelligence

Chair: Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara

1 - Algorithmic Bidding in Procurement Auctions

Jingyi Tian, University of Houston, Houston, TX, United States

As online auction platforms evolve rapidly, they amass vast data from repeated interactions, prompting bidders to harness AI technology for strategic advantage. This paper explores the intriguing role of AI bidders in online procurement auctions, driven by this technological shift. Through a novel analytical model, we demonstrate how the Deep Q-Network (DQN), a widely used reinforcement learning algorithm, achieves a stable auction equilibrium. Our findings reveal a fascinating twist: the algorithm, operating with incomplete information, learns a correlated objective that leads to overbidding. We further investigate market conditions that enable collusive algorithmic overbidding through the DQN updating process. Complementing our analytical insights, we also conduct experiments with AI bidders to empirically confirm stationary convergence on algorithmic overbidding and assess measures to mitigate this inefficiency. Our results highlight the peril of misspecified correlated objectives in AI-driven multi-agent environments, showing stark contrasts in decision-making and equilibrium outcomes compared to rational human bidders. In light of this discrepancy, we offer strategic recommendations for refining procurement processes in auctions to minimize algorithmic overbidding. Additionally, we discuss broader implications for market design on digital platforms, advocating for human-centered approaches amid the rapid surge in AI adoption. Our work underscores the critical need for thoughtful AI integration in auction markets, balancing technological advancements with effective governance.

2 - Target Collection using Efficient Reinforcement Learning Algorithms via Posterior Sampling for Distributionally Environments

M. Amin Rahimian, University of Pittsburgh, Pittsburgh, PA, United States, CARLOS HURTADO, Lei Fang, Ze-Xu Li

Consider a soldier/robot(Agent) that is searching for targets in a bounded region (Environment). The agent does not have information on the exact location of the targets, but that their locations are sampled from a given distribution. Target locations are sampled from a hierarchical normal model with two levels. The first level uses an overall mean and overall variance-covariance matrix to sample centers of patches. Therefore, we use each center of the patch as a mean parameter to sample targets per patch. We want to find a set of movement directions, also known as policy, such that we maximize the number of targets collected and minimize the total distance traveled. This optimization problem is equivalent to maximizing the search efficiency which is the ratio between targets collected and distance traveled. We take advantage of the structure of the target distribution to design a policy while the agent interacts with the environment and collects more information. Thus, we use the following key ideas to implement a policy as a solution to our problem. If the targets' locations are known, this problem can be stated as a shortest path problem. Which is easy to solve, where each target location is a node to visit in the shortest path problem. Either if the target is found or not, we can update the overall hierarchical distribution model. We show that this approach learns the optimal path between nodes, and we provide convergence rates as a function of the parameters of the hierarchical distribution.

3 - AI Spillover Impact on Real Estate Employment

lingyu li, Virginia Commonwealth University, Richmond, VA, United States, Yeongin Kim, Xiaojin Liu

This study explores the AI technology spillover effects. This study investigates the spillover impact of AI on employment in the real estate industry. We model AI technology using the upgrades of Zestimate in Zillow. We explore how the upgrade of Zestimate in Zillow spillover to affect the employment of real estate agents. In this paper, we apply a difference-in-difference and propensity score matching method. We find a significant negative spillover impact of AI technology application on employment in the real estate industry. These impacts are more pronounced in poorer areas and high technology penetration areas. Our work has practical implications for policymakers in formulating AI technology-related policies. Furthermore, our study supports the argument that AI technology helps alleviate the imbalance between the poor and the rich, thereby improving overall social welfare.

4 - AI Driven Software Development: An Empirical Analysis

Karthik Babu Nattamai Kannan, Southern Methodist University, Dallas, TX, United States, Narayan Ramasubbu

Generative AI is significantly transforming the landscape of software development. By leveraging advanced machine learning algorithms, GenAI automates various aspects of the development process, thereby increasing efficiency, reducing errors, and fostering innovation. These tools help developers generate code, summarize their changes, test and debug the code. Prior research points to an increase in developer productivity due to GenAI tools. In this paper, we examine how the introduction of GenAI tools impacts the software development. We exploit a quasi-experimental setting in which GenAI tools were introduced during certain time periods of the development process. Using a difference-in-differences framework, empirically examine how the adoption of these GenAI tools impacts the different aspects of the software development process.

TE67

Regency - 605

AI Innovations: Text Interaction to Trustworthy Decisions

Contributed Session

Chair: Petros Xanthopoulos, Stetson University, DeLand, FL, United States

1 - Committee Configuration Optimization for Parallel Consensus

Yifei Xie, University of Edinburgh, Edinburgh, United Kingdom

Current blockchain applications encounter significant scalability challenges. A practical solution to this issue involves partitioning the consensus network into multiple committees and executing consensus processes in parallel, a technique known as sharding. However, achieving optimal partitioning in distributed environments is complex due to potential node failures and network communication delays.

To address this problem, we propose a stochastic programming method designed to enhance the performance of parallel consensus in sharded blockchain networks. Our primary objective is to develop an optimal partition configuration that maximizes performance in the sharded blockchain network, accounting for uncertain network conditions and dynamic system states. This approach integrates the strengths of traditional methods, random configuration, and mathematical programming. We applied stochastic programming to optimize both the replication protocol during normal operations and the view change protocol in the event of a leader node failure.

To evaluate the effectiveness of our method, we conducted experiments in a realistic testing environment using Microsoft Azure cloud virtual machines. Our results demonstrate a consistent performance improvement of approximately 24% in parallel consensus, thereby validating the efficacy of our proposed method.

2 - A text similarity based tool against AI-assisted paraphrasing

Petros Xanthopoulos, Stetson University, DeLand, FL, United States, Konstantinos Xylogiannopoulos, Panagiotis Karampelas, Georgios Bakamitsos

Large Language models (LLMs) can be used to paraphrase human text that in turn can be used in a number of malicious AI-assisted plagiarism applications. Unlike conventional AI text detection, paraphrased text can fly under the radar of commercial AI detection classification tools. In this presentation we demonstrate certain similarity based patterns that can be used for the successful detection of AI plagiarized text generated by ChatGPT. We demonstrate the application of this feature into a number of application such as fake online review detection and fake news articles detection.

3 - Impacts of Resource Sharing in Smart Manufacturing on Operations Strategy: A Text Mining Approach

Ai-Hsuan Chiang, Ming Chuan University, Taipei, Taiwan

With the advent of Industry 4.0, smart manufacturing has become a trend in the manufacturing industry. Smart manufacturing enhances production flexibility to create customized products and reduces the need for labor, enabling enterprises to lower various costs. This article focuses on the relationship between smart manufacturing and operations strategies. Previous research indicates that sharing and utilizing resources can bring greater benefits to enterprises. Therefore, this study explores the impact on competitive priorities when enterprises implement knowledge sharing and physical resource sharing under smart manufacturing. This research employs text mining methods, using corporate annual reports and data from the Taiwan Economic Journal to analyze research model from 281 electronic manufacturing companies. The results show that smart manufacturing significantly influences both cost-first competitive priority and customization-first competitive priority. Knowledge sharing positively affects the choice of a cost-first competitive priority for enterprises. If an enterprise opts for a customization-first competitive priority, it needs to implement both physical resource sharing and knowledge sharing. The findings illustrate that under smart manufacturing, the benefits can only be maximized when the chosen strategies and resource-sharing behaviors of an enterprise are aligned.

4 - Do Not Let the Lexicons and Their Components Drive Your Results

Eduardo Villacis-Calderon, University of Texas at El Paso, El Paso, TX, United States, Sergio Grove, Yoonsun Jeong

Given the popularity and widespread application of computer-aided text analysis (CATA) approaches in lexicon-based text analysis, we present a novel issue that researchers usually overlook. Our research examines the implications of improper use of difference scores in sentiment analysis paired with whimsical lexicon selection criteria. We replicate prior research on perceived review usefulness that used difference scores to measure overall sentiment (positive minus negative sentiment) as the sole sentiment construct. Our study demonstrates that inadequate approaches to using overall sentiment scores may result in empirical bias that can lead to opposite effects depending on the lexicon employed. We raise awareness of substantial methodology weaknesses that affect researchers in their theorization and testing. Using a sample of 21,016 consumer reviews from Amazon.com, our findings call for changes in methodology when using the overall sentiment score and selecting lexicons. Accordingly, we provide practical recommendations and minimum testing standards for CATA lexicon-driven models in business research.

TE68

Regency - 606

Innovative Approaches in AI, Supply Chains, and Multi-Agent Systems

Contributed Session

Chair: Kyeonghyeon Park, KAIST, 291, Daehak-ro, Yuseong-gu, Daejeon, 34141, Korea, Republic of

1 - Hypernetwork-based approach for optimal composition design in partially controlled multi-agent systems

Kyeonghyeon Park, KAIST, Daejeon, Korea, Republic of, David Molina Concha, Hyun-Rok Lee, Chi-Guhn Lee, Taesik Lee

A Partially Controlled Multi-Agent System (PCMAS) is comprised of two types of agents: those that are controllable, directly managed by a system designer, and those that are uncontrollable, operating autonomously. In scenarios where agents' autonomy must be preserved, effectively managing controllable agents is essential for steering the system towards desired outcomes. In this study, we investigate a new approach to the composition design problem faced by the system designer in PCMAS, wherein the optimal number of controllable agents within a budget constraint should be found. Besides determining the optimal number of controllable agents, this problem also involves crafting cooperative policies for the controllable agents and identifying the best response policies for the uncontrollable agents. Obtaining the solution to this problem requires repeatedly solving multi-agent reinforcement learning (MARL) for various configurations, which is computationally demanding. To overcome these obstacles, we introduce a hypernetwork-based method that optimizes system composition and cooperative policies. Instead of training independent policy networks for each composition, we utilize a hypernetwork architecture that generates the weights of another neural network, enabling unified training across all system compositions. Through this approach, we generate policies for both controllable and uncontrollable agents across various system compositions, mitigating computational burden and lack of information sharing in similar compositions. We assess how closely the generated policies approximate equilibrium policies and demonstrate performance improvements resulting from optimized system composition facilitated by the trained hypernetwork.

2 - Modeling human behavior with incomplete data; a hybrid approach integrating agent-based modeling and machine learning

Ali Mahdavi Adeli, University of Memphis, Memphis, TN, United States

To create simulation models of socioeconomic phenomena (e.g., human behavior in financial markets), we can think of two extremes on a continuum. On one extreme, we have an established theory or full understanding of the phenomena (and/or its domain) and can therefore hardcode (stochastic or deterministic) decision rules to create, for example, agent-based models, or specify analytical models with certain simplifications and assumption about the domain (e.g., Cobb-Douglas production function in economics). On the other end, we might not have an understanding or theory of the phenomena but have large amounts of data, which allow us to use data-hungry machine learning (ML) or deep learning approaches to create models that replicate the observed phenomena or predict outcomes of interest. In between these two extremes, we have situations where there is partial understanding about the phenomena and partially observed data (on outcomes and variables of interest), but not enough to successfully utilize purely ML based approaches. We propose a hybrid approach integrating ML techniques with agent-based modeling to model human behaviors in such scenarios, i.e., where there is partial understanding and partially observed data available. We provide a guideline to determine the suitability of our proposed approach given a problem scenario and demonstrate its effectiveness by creating a simulation model of human decision making in the complex decision environment of combinatorial auctions. We show how our approach provides, otherwise unfeasible, insights on auction outcomes and participants' strategies under all plausible scenarios.

3 - The Limits of Static Decision Rules in Supply Chain Games

el ouardighi fouad, essec business school, cergy, France

This talk suggests examples of supply chain game models, each associated with specific scenarios involving one manufacturer and either one supplier or one retailer. Both shortsighted and farsighted decision rules are derived from the models. Our analysis exposes the underlying mechanisms contributing to the suboptimality of static policies compared to their dynamic counterparts.

4 - Achieving Sustainable Supply Chains through Artificial Intelligence: Systematic Literature Review

Abdullah Albizri, Montclair State University, Montclair, NJ, United States, Marina Evy Johnson, Antoine Harfouche

The sustainability and supply chain management research studies have become more theoretically rich and methodologically rigorous. Similarly, there have been many significant advances in Artificial intelligence (AI) and an increase in its adoption that AI-based technologies drive many industry functions and operations. Therefore, there is an increasing demand for investigating the ways to create and manage sustainable supply chains, particularly through AI. To this extent, this paper provides a systematic, rigorous, and methodologically valid review of sustainable supply chain management and AI literature to extract insights and identify AI-related key factors to create and manage sustainable supply chains. This study also determines the most recent trends and gaps to develop a new research agenda to guide future sustainable supply chain management studies.

TE69

Regency - 607

Sports IV

Contributed Session

spORts

Chair: Timothy Kaskela, Oregon State University, Corvallis, OR, United States

1 - Game Theory and Sports Analytics with Applications to Baseball

David Grimsman, Brigham Young University, Provo, UT, United States, Christopher Archibald, William Melville

Game theory is a mathematical tool used to analyze strategic decision-making in complex environments. Its application to sports is a natural fit, given the recent abundance of data and well-defined rules and outcomes. In this talk we will give a brief overview of game theory and sports analytics, including problem formulation and the fundamental tradeoff between optimization and robustness. We will discuss a recent line of work in baseball: an analysis of the batter-defense interaction as a zero-sum game. The work presents a methodology for choosing an optimal pitch sequence and fielder positioning by solving for a game equilibrium that is a generalization of Nash equilibrium and Stackelberg equilibrium. The presentation will conclude with some possible game-theoretic problem formulations and expected challenges in other major sports.

2 - Teaching and Engaging Students with Sports Data: SCORE National Network

Andrew Lee, US Military Academy, West Point, NY, United States

Many students are athletes and/or avid sports fans who appreciate and find motivation in applications of statistics to issues that arise in sports. Where can instructors (or students) find relevant and accessible sports data? What are some sports questions that map well to typical course objectives? How can we structure sports-based examples to be accessible and interesting to non-fans? How can instructors publish their own instructional modules? I will share information about the SCORE network, which is a national network for developing and disseminating Sports Content for Outreach, Research, and Education in data science, funded through the NSF. The mission of SCORE is to create a sustainable national network of academics, industry and government partners that will build the SCORE platform through which we will develop and ensure the sustainability of our efforts in elevating statistics and data science education, particularly in underrepresented populations and minorities.

3 - Who is the GOAT? Sports Rankings and Random Walks on the Symmetric Group

Juan Martinez Mori, Georgia Institute of Technology, Atlanta, GA, United States, Gian-Gabriel Garcia

Given a collection of historical sports rankings, can one determine which athlete is the greatest of all time (i.e., the GOAT)? In this work, we introduce a data-driven random walk on the symmetric group to obtain a stationary distribution over athlete rankings, spanning across different time periods in sports history. We combine this distribution with a notion of stochastic dominance to obtain a partial order over the athletes. We implement our methods using publicly available data from the Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA) to find the GOATs in the respective categories.

4 - Personalized Endurance Training: Leveraging Data for Runner Optimization

Marc De Reu, University Ghent (Belgium), Gent, Belgium, Dirk Van den Poel

In recent years, training theory for endurance sports has emphasized optimizing training loads. Not only elite runners but also recreational marathon runners rely on personalized training plans to improve their performance and prevent injuries. Previously, only experienced human coaches could design such plans. However, we develop algorithms that use detailed training and racing data to generate customized training plans for each runner.

Our model uses a comprehensive dataset with specific details of training sessions and race performance to tailor each training session to the runner's unique profile, time available to train and performance goals. Our machine-learning model analyses large amounts of data to adjust training variables such as intensity and volume, minimizing the training load for each runner, and providing feedback to both the runner and the coach.

Our approach has shown promising results, such as an improvement in performance by approximately 2% - 3% and a decrease in injuries by more than 50% compared to traditional training methods. While these initial results are promising, we believe that including additional lifestyle factors such as sleep patterns, stress levels and dietary habits could further refine the accuracy of our training algorithms. Our goal is to develop even more personalized and effective training plans that take these broader personal lifestyle variables into account.

Our new approach demonstrates the potential of using detailed data in sports training and paves the way for more advanced adaptations in personalized coaching in endurance sports.

5 - Nfl Referee Bias Based on Perceived Differences in Quality and Score Difference

Timothy Kaskela, Oregon State University, Corvallis, OR, United States

Research has consistently found that referees are affected by bias in numerous contexts. Prior research has focused on how bias occurs against each team, finding bias based on factors such as home team and distance of fans to the playing area. This research examines how penalties are called based on a different context: how penalties are called on the winning team. As both teams could be winning at different times during the game, the examination of penalties does not solely focus on one team or the other. Instead, I examine how penalties are called on the winning team at that specific point in the game. This examination is conducted through play-by-play data for every NFL game from 2016-2022. As I focused on plays in which a penalty was called, the dataset was narrowed down to only those plays that resulted in a penalty, for a final total dataset of 21,473 observations. Moderating factors such as the difference in perceived quality of each team, score differential, and subjectivity of a penalty are taken into account. I find that there is an inverse u-shaped relationship between score differential and probability of a penalty being called on the winning team and the difference in perceived quality of the teams has a significant effect on said relationship. In addition, while prior research has found that the number of penalties called on each team tends to be fairly balanced, I find that penalties when split between when a team is winning or losing is unbalanced.

TE70

Regency - 701

Latest Practices of Stochastic and Robust Optimization for Renewable Energy

Invited Session

ENRE: Electricity

Chair: Guangchun (Grant) Ruan, MIT, Boston, MA, United States

1 - Planning and Investment in Electricity Markets with Abundant Penetration of Offshore Wind Using Quantile Fourier Regressions and Markov Decision Processes

Golbon Zakeri, University of Massachusetts - Amherst, Amherst, MA, United States, Geoffrey Pritchard, Arash Khojaste

We will present a model for daily operation of electricity markets with abundant penetration of renewable resources of electricity generation, e.g. offshore wind. Our model captures the underlying uncertainty based on inhomogeneous Markov models whose states are determined from quantiles of the underlying data. To incorporate seasonal and diurnal variations smoothly and have a simple parametric form, we appeal to Fourier basis functions.

2 - Designing Renewable Power Purchase Agreements: Impact on Green Energy Investment

Zuguang Gao, University of California, Irvine, Irvine, CA, United States, Nur Sunar, John Birge

We study a firm that aims to sign a power purchase agreement (PPA) with a renewable energy producer. Using a stochastic control framework, we identify the optimal renewable PPA for the firm. Our analysis generates various valuable insights for both managers and policy makers.

3 - Optimization for Data-Driven Modeling and Control of Inverter-Based Resources

Javad Khazaei, Lehigh University, Bethlehem, PA, United States

A complex physics-based modeling procedure and the uncertainty and confidentiality of internal parameters of cyber-physical systems motivate data-driven system identification tools. With the availability of high-fidelity measurements and historical data, model-free identification of nonlinear dynamical systems can facilitate the control design without tedious physics-based modeling procedures. An application domain that can benefit from such exploration is the electricity grid with many inverter-dominated cyber-physical systems such as renewables, energy storage, electric vehicles, and various nonlinear loads, which is complex to model using conventional approaches. This presentation introduces statistical learning and sequential optimization techniques for data-driven nonlinear modeling and control of distributed energy resources (DERs) in smart grids using sparse identification. Using the proposed data-driven model identification tool for closed-loop control of DERs, we demonstrate the effectiveness of a model-free design in successful control and stability analysis of nonlinear DERs in smart grids. Compared with existing physics-based designs that heavily rely on knowing the detailed system dynamics or uninterpretable data-driven designs that rely on large historical data, the proposed model-free DER identification and control frameworks can accurately capture the dynamics of the DERs based on available measurements and provide guaranteed performance for black-start, weak AC grid integration, microgrid integration, and stability analysis.

4 - Distributionally Robust Chance-Constrained Optimal Building HVAC Control for Renewable Integration

Jin Dong, Oak Ridge National Laboratory, Oak Ridge, TN, United States, Yiling Zhang

Aggregation of heating, ventilation, and air conditioning (HVAC) loads can provide reserves to absorb volatile renewable energy, especially solar photo-voltaic (PV) generation. Given the uncertain and time-varying nature of solar PV generation, its probability distribution is difficult to be estimated perfectly, which poses a challenging problem of how to optimally schedule a fleet of HVAC loads to consume as much as local PV generation. We formulate a distributionally robust chance-constrained (DRCC) model to ensure that PV generation is consumed with a desired probability for a family of probability distributions under two typical ambiguity sets: the moment-based and Wasserstein ambiguity sets. We derive mixed integer linear programming (MILP) reformulations for DRCC problems under both sets.

TE71

Regency - 702

AI and Data Science for Smart Systems

Invited Session

Data Mining

Chair: Jihoon Chung, Pusan National University, Busan

Co-Chair: Emmanuel Yangue, Oklahoma State University, Stillwater, OK, United States

1 - Physics Informed Neural Network Using Sparse Identification of Differential Equation with Application to Additive Manufacturing

Jihoon Chung, Pusan National University, Korea, Korea, Republic of, Raghav Gnanasambandam, Bo Shen

Differential equations are fundamental in modeling numerous physical systems, including thermal, manufacturing, and meteorological systems. Due to the development of deep learning, Physics-informed Neural Networks (PINN) are evolving as a new paradigm for solving differential equations. However, most previous work assumes that differential equations are given, which is an unrealistic assumption in practice. Therefore, we develop the framework to determine the accurate differential equations for PINN. Specifically, we focus on the sparsity of possible functions of the differential equation to derive accurate equations according to data. The actual additive manufacturing data is utilized to validate the proposed method's effectiveness.

2 - Noise Conditioned Diffusion Denoising Implicit Model (Ddim) with Mixed Sampling Distribution and Adversarial Discrimination

Emmanuel Yangue, Oklahoma State University, Stillwater, OK, United States, Chenang Liu, Wenmeng Tian, Feng Yu

Given the rapid progress in deep learning, numerous studies have employed its capacity to tackle diverse challenges within specific fields. Within engineering applications (e.g. additive manufacturing (AM) and disease treatment using medical imaging), the analysis and interpretation of images are pivotal for advancing system capabilities and expanding their utility. This study introduces an advanced diffusion-based generative model with adversarial capability. In this novel approach, the diffusion model employed, a denoising diffusion implicit model (DDIM), serves as the generator in the adversarial framework to produce high-quality synthetic images. Within this model, a variational encoder conditions the diffused noise injected into input images with prior information from the training images. Afterward, the generated images go through a timestep discriminator to assess their realism. The noise distribution is based on a hybrid sampling technique that combines gamma and Gaussian distributions. The proposed model underwent testing across various case studies, including layer-wise AM images and tumor ultrasound medical images. Current findings demonstrate the substantial potential of this approach to provide solutions to real-world engineering applications.

3 - Multi-Objective Bayesian Optimization Framework for Extreme Value Analysis

Shehzaib Irfan, Auburn University, Auburn, AL, United States, Jia Liu

Extreme value theory (EVT) can model and predict extreme events, such as floods, heatwaves, and financial crises. The block maxima (BM) sampling is usually used to divide the region of interest (e.g., years) into blocks and extract the maximal values from each block for EVT analysis. However, BM lacks a mechanism to propose an optimal block size for extracting the maximal values. This work proposes a Bayesian optimization framework for finding the optimal block size for modeling the extremes. The framework defines two objective functions that balance goodness-of-fit for distribution of maximal values and prediction accuracy for extreme values. The optimal block size is found from a Pareto frontier by using Multi-Objective Bayesian Optimization. In the case study, using the optimal block size from the proposed framework for EVT analysis has achieved an improvement of about 50% in prediction accuracy. Such framework can be extended to different applications when using EVT to analyze extreme events.

4 - Personalized Federated Learning for Enhanced Predictive Modeling in Additive Manufacturing

Rong Lei, Rutgers University-New Brunswick, Piscataway, NJ, United States, Yuebin Guo, Weihong Guo

Additive Manufacturing (AM) processes can significantly benefit from process monitoring, anomaly detection, and defect prediction. Current AM processes are characterized by significant variability and complexity, often leading to instability and defects. Averaging-based federated learning (FL) approaches, designed for collaborative learning without data sharing, struggle with the inherent heterogeneity of AM data. This paper introduces a novel knowledge-distillation based FL framework tailored for the AM industry, particularly addressing small- to medium-sized enterprises (SMEs) that crowdsource shared resources. The proposed framework incorporates personalized models that leverage local knowledge and dynamic distillation techniques to improve predictive accuracy and robustness across diverse AM scenarios. Our approach enhances data preprocessing to reduce heterogeneity and accelerates model convergence. We design a dynamic evaluation of knowledge transfer to balance global model accuracy with local personalization. Additionally, the framework incorporates customized aggregation indices to guide the selection of local training models and hyperparameters.

TE72

Regency - 703

Advances in Quantitative Methods for Business Analytics

Invited Session

Data Mining

Chair: Jessica Leung, Monash University, Melbourne, 3145, Australia

1 - High-Dimensional Visualisation via Projection Pursuit: Performance of the jellyfish optimizer

Jessica Leung, Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia, Sherry Zhang, Nicolas Langrené, Di Cook

This study explores the integration of the jellyfish algorithm, a metaheuristic optimizer, in enhancing high-dimensional visualization through the projection pursuit guided tour. The jellyfish algorithm, inspired by the natural foraging behavior of jellyfish, is utilized for its efficiency in navigating the complex search spaces typical of high-dimensional datasets. Through extensive simulations, our approach demonstrates improved efficacy in uncovering underlying data structures that are not readily apparent. Our findings suggest that using the jellyfish algorithm markedly enhances the discovery of insightful projections, facilitating a deeper understanding of multidimensional data landscapes.

2 - Market Timing with Bi-objective Cost-sensitive Machine Learning

Robert James, The University of Sydney, Sydney, Australia, Jessica Leung, Artem Prokhorov

This paper develops a framework for cost-sensitive training of machine learning models that predict the direction of aggregate stock returns. We design a bi-objective loss function that augments the traditional log-loss objective with an objective that minimizes the cost of individual false-positive and false-negative classification errors. We argue that the option implied tail conditional value-at-risk is a natural measure of misclassification costs. Our bi-objective optimization framework permits us to isolate the effect of cost-sensitivity from log-loss minimization, and integrate forward-looking information from options markets directly into the model training process. We study changes in the classification behavior of elastic-net logistic regression and gradient boosted decision trees trained using our bi-objective loss function. These new models improve the risk-adjusted returns of market timing strategies and substantially reduce downside risk.

3 - TBA

4 - Enhancing cyber resilience of crown jewels in cyber supply chains

Kam-Fung Cheung, University of New South Wales Business School, Sydney, Australia, Chung-Li Tseng, Michael Bell

The COVID-19 pandemic has accelerated the digitisation of businesses, while its associated cyber threats affect various industries such as logistics and shipping. The unavailable crown jewels, i.e., critical assets, due to successful cyberattacks undermine the focal organisation's and its partners' profits, thus there is an urgency in enhancing cybersecurity and cyber resilience of critical assets in supply chains. This work proposes a novel integer programming model based on eigenvector centrality to identify (hidden) critical assets in a cyber supply chain network and improve their security by implementing effective cybersecurity measures. Managerial insights for enhancing cybersecurity and cyber resilience for critical assets are provided to help develop defensive strategies.

5 - On Improved Semi-parametric Bounds for Tail Probability and Expected Loss

Artem Prokhorov, University of Sydney Business School, CEBA, CIREQ, Sydney, Australia, Erick Li

We revisit the fundamental issue of tail behavior of accumulated random realizations when individual realizations are independent, and we develop new sharper bounds on the tail probability and expected linear loss. The underlying distribution is semi-parametric in the sense that it remains unrestricted other than the assumed mean and variance. Our sharp bounds complement well-established results in the literature, including those based on aggregation, which often fail to take full account of independence and use less elegant proofs. New insights include a proof that in the non-identical case, the distributions attaining the bounds have the equal range property, and that the impact of each random variable on the expected value of the sum can be isolated using an extension of the Korkine identity. We show that the new bounds not only complement the extant results but also open up abundant practical applications, including improved pricing of product bundles, more precise option pricing, more efficient insurance design, and better inventory management.

TE73

Regency - 704

Data Analytics and Machine Learning in Healthcare

Invited Session

Data Mining

Chair: Tianshu Feng, George Mason University, Fairfax, VA, United States

1 - Variational and Explanatory Neural Networks for Encoding Cancer Profiles and Predicting Drug Responses

Tianshu Feng, George Mason University, Fairfax, VA, United States, Rohan Gnanaolivu, Yuanhang Liu, Jun Jiang, Alexander Partin, Priyanka Vasanthakumari, Yitan Zhu, Chen Wang

Human cancers present a significant public health challenge, necessitating novel drug discoveries through translational research. Transcriptomics profiling data, detailing molecular activities in tumors and cancer cell lines, are widely utilized for predicting anti-cancer drug responses. De-spite previous efforts to develop data-driven AI/neural network (NN) models for cancer studies, their effectiveness is impeded by a lack of robust encoders to account for noise in continuously measured transcriptomics data and difficulty achieving biological interpretability in sophisti-cated learning tasks, such as drug response predictions. To address these limitations, we intro-duce a novel neural network framework, VETE (Variational and Explanatory Transcriptomics Encoder), which incorporates a variational component to combat noise effects and integrates traceable gene ontology into NN architecture to encode cancer transcriptomics data. Key innova-tions include a local interpretability-guided method for identifying connected ontology terms as an interpretable path for explaining predictions and an interactive visualization tool to elucidate likely biological mechanisms for different anti-cancer drug responses. The VETE framework demonstrated robust classification accuracies in patient tumors and cancer cell lines. Moreover, we investigated the interpretability of VETE in classifying human cancer cell-lines according to tissue of origins and predicting drug responses, providing traceable biological explanations. VETE represents a promising advancement in the field, bridging the gap between AI-driven pre-dictions and biologically meaningful insights in cancer research.

2 - Artificial Intelligence for Nano Particle Folding Prediction of Rna

You Zhou, Arizona State University, Tempe, AZ, United States

The approaches to predict the RNA secondary structure were mainly realized by minimizing the free energy previously. Due to the large success of utilizing Deep Learning to predict the protein structure, utilizing Deep Learning becomes a new direction to handle RNA structure prediction. In this study, we propose a novel Deep Learning approach (Transformerfold) to predict the RNA secondary structure. The Transformerfold is based on Transformer whose predictions are reinforced by our previously proposed RNA sequence-secondary structure pair evaluation models, NU-ResNet and NUMO-ResNet. The Transformerfold with its open-source software would be beneficial to RNA folding field.

3 - SAM-Driven Explainable and Uncertainty-Aware Automatic Measurement Model in Echocardiography

Beom Gyo Shin, Yonsei University, Seoul, Korea, Republic of, Jaeik Jeon, Hyunsoo Yoon

In echocardiography, an estimation of left ventricular (LV) volume plays a key role in assessing cardiac diseases, yet it poses challenges. Previous research has relied on segmentation-based approaches, which incur significant costs and time for labeling. Additionally, in quantitative medical image analysis, there exists inter-intra observer variability. This study introduces a novel probabilistic deep learning framework for measuring the LV volume and quantifying associated uncertainty, without relying on segmentation labels. Furthermore, the proposed method maintains the explainability of the model by generating LV boundary using weak supervision techniques and Segment Anything Model (SAM). Experimental results demonstrate competitive performance in predicting both LV volume and boundaries and estimating uncertainty.

4 - Inferring Drug-Induced Plasticity via Drug Screen Data

chenyu wu, University of Minnesota - Industrial and Systems Engineering, MINNEAPOLIS, MN, United States, Kevin Leder, Jasmine Foo

Resistance to drugs remains a significant challenge in cancer treatment, often attributed to the presence of a cancerous stem-like subpopulation that drives tumor recurrence post-treatment. Compounding this issue, many drugs targeting cancer cells inadvertently induce plasticity, reverting differentiated cancer cells to a stem-like state. Addressing this phenomenon is crucial for advancing cancer therapeutics. In this study, we introduce a robust statistical framework founded in multi-type branching processes to dissect both drug effects and tumor dynamics using high throughput screening data. Through comprehensive *in silico* experiments, we demonstrate the efficacy of our framework in accurately estimating parameters governing population dynamics and drug responses. Moreover, our framework is applied to recent *in vitro* data, spotlighting the drug-induced plasticity of ciclopirox olamine in gastric cancer cells. Leveraging established model selection criteria, our analysis yields consistent insights into the impact of ciclopirox olamine on gastric cancer cell behavior.

5 - Modelling hospital bed and ventilator capacity during COVID-19

Ozlem Cosgun, Montclair State University, Montclair, NJ, United States, Eyyub Kibis

During the COVID-19 outbreak, all countries faced a critical shortage of beds and ventilators that caused a dramatical increase in mortality rates. In this study, we addressed the optimization problem of allocating bed and ventilator capacities by explicitly incorporating Non-Pharmaceutical Interventions (NPIs) into an extensive SIR model to minimize the number of deaths in New York City during COVID-19. A mixed integer programming model was developed to decide on the optimal bed and ventilator expansion amounts that minimize the number of deaths over time.

TE74

Regency - 705

Heat Decarbonization

Invited Session

ENRE: Energy-Climate

Chair: Jacqueline Dowling, Carnegie Science, 260 Panama St., Providence, 94305, United States

Co-Chair: Henry Ssembatya, NC State University, RALEIGH, NC, United States

Co-Chair: Aleksander Grochowicz, University of Oslo, Oslo, Norway

1 - The Role of Policy and Module Manufacturing Learning in Industrial Decarbonization by Small Modular Reactors

Max Vanatta, University of Michigan, Ann Arbor, MI, United States, Michael Craig, William Stewart

Small nuclear modular reactors (SMRs) offer a unique solution to the challenge of decarbonizing mid- and high-temperature industrial processes. We develop deployment pathways for four SMR designs serving industrial heat processes at 925 facilities across the United States under diverse policy and factory and onsite learning conditions. We find that widespread SMR deployment in industry requires natural gas prices above $6/MMBtu or aggressive carbon taxes. At natural gas prices of $6 to $10/MMBtu, 7 to 55 GWt of SMRs could be economically deployed by 2050, reducing annual emissions by up to 59 million mt CO2e. Widespread deployment potential hinges on avoiding substantial cost escalation for early investments. Policy levers like direct subsidies are not effective at incentivizing sustainable deployment, but aggressive carbon taxes and investment tax credits provide effective support for SMR success. Large-scale SMR deployment hinges on factory, not onsite, learning.

2 - Evaluating Tradeoffs of Residential HVAC Electrification in a Changing Climate

Parth Vaishnav, University of Michigan, Ann Arbor, MI, United States, Shuhaib Nawawi, Michael Craig, Ming Yi

Decarbonizing the residential sector requires the electrification of the heating, ventilation, and air conditioning (HVAC) systems, with air source heat pumps (ASHP) emerging as a promising technology. To plan for this transition, analysts must evaluate the effect of widespread ASHP adoption at an appropriate temporal resolution, focusing on both household and power grid. In response, we develop a reduced complexity building energy model (RC-BEM) to predict the hourly energy demand of over 17k statistically representative single-family homes across 54 U.S. cities. Using RC-BEM, we co-optimize occupant- and grid-related objectives using historical and future climates. Our trade-off analysis reveals three key findings: i) energy cost dominates the optimization and is the most sensitive to different comfort levels, followed by monetized human health and environmental damages from electricity consumption, and peak demand cost, ii) conventional HVAC sizing strategies based on historical climate may be inadequate for ensuring thermal comfort in a warmer climate, and iii) using the price of electricity as a proxy for thermal discomfort is likely to underestimate the true disutility experienced.

3 - A Comparison of Integrated Power System and Natural Gas Planning under Building Sector Decarbonization Policies

Stephanie Wilcox, Johns Hopkins University, Baltimore, MD, United States, Ben Hobbs, Ozge Kaplan

Policies aimed at decarbonizing the building sector seek to mitigate greenhouse gas (GHG) by promoting the use of cleaner technologies, restricting access to GHG-intensive fuels, or transitioning electricity generation towards lower carbon-sources, for example. However, the economic implications, operational effects, and investment impacts of these policies are often understudied. Our research, using a generation and expansion planning model and natural gas distribution system steady-state flow model, indicates that power system decarbonization goals result in minimal building sector GHG reductions. Our findings also suggest that the most cost-effective GHG-reducing solutions consider homogeneity of building archetypes and local natural gas and power system network characteristics, such as capacity and age.

4 - Stochastic Power-Gas Infrastructure Planning to Meet the Challenges of Heating Electrification

Rahman Khorramfar, MIT, Cambridge, MA, United States, Dharik Mallapragada, Saurabh Amin

Central to economy-wide decarbonization efforts is building heat electrification which directly impacts power-gas infrastructure through replacing gas with electric power. Here, we develop a modeling framework to quantify the end-use demand for both power and gas in the residential building sector under various electrification pathways. We then evaluate their impact on the joint power-gas investment and operational planning under uncertainty to ensure that the future grid is resilient to interannual weather variations. This modeling framework is motivated by the uncertain impacts of climate change on renewable energy potential and shifting demand patterns in many socio-economically vulnerable regions. We propose two mathematical paradigms, namely stochastic programming (SP) and distributionally robust optimization (DRO), to characterize the underlying uncertainty in power-gas demand as well as generation potential of renewables. We develop a computational approach that exploits the spatial correlation between electric power and natural gas demands both within and across load zones, resulting in a mixed-integer programming (MILP) formulation. We compare the resulting uncertainty-aware designs with the certainty equivalent problem for a realistic power-gas system in the U.S. New England region and highlight the importance of systematic modeling of inter-annual variability in demand and renewable energy supply.

TE75

Regency - 706

Operational and Data-Centric Perspectives on Sustainability

Invited Session

ENRE: Environment and Sustainability

Chair: Ilgin Dogan, INSEAD, Fontainebleau, 94707, France

Co-Chair: Ho-Yin Mak, Georgetown University, Washington, DC, United States

1 - Data-Driven, Large-Scale Inventory Routing for Energy Solutions at the Bottom of the Pyramid

Dan Iancu, Stanford University, Stanford, CA, United States, Sergio Camelo, Maximilian Schiffer, Simon Thoma

We propose a large-scale inventory routing problem in the context of distributing clean cooking stoves in East Africa. Using a comprehensive dataset from an industry partner, we develop a stochastic model that accounts for significant supply and demand fluctuations, censored and incomplete data, and propose a data-driven approach to solve the resulting problem at scale. Our numerical results demonstrate that the proposed method yields significant improvements over the status quo, leading to significant cost savings and improved accessibility to clean cooking solutions.

2 - Exploring the Distribution Dynamics of Clean Cooking Solutions

Banu Tiryaki, Stanford University, Stanford, CA, United States, Dan Iancu, Omer Karaduman

This study examines the role of third-party service providers in the distribution and adoption of clean cooking solutions in Kenya. The emergence of new clean cooking solutions has created job opportunities that not only facilitate the adoption of these innovations but also potentially support the businesses providing them. Our research models the operations of intermediaries connecting users with businesses and examines individual participation in these third-party services. By exploring the dynamics of these intermediaries, we aim to understand their impact on the market and the businesses supplying clean cooking solutions.

3 - Planning Stormwater Retention for Resilience Against Extreme Rainfalls

Aiqi Zhang, Wilfrid Laurier University, Waterloo, ON, Canada, Sheng Liu, Wei Qi

Climate changes inflict prolonged and intensified rainfalls on cities around the world. Unfortunately, existing efforts hardly meet the urgent need for climate-adaptive stormwater management. Urban stormwater infrastructure is typically planned based on empirically predetermined rainfall scenarios on the intensity-duration-frequency (IDF) curve, which fails to capture the worst-case scenarios that should be planned against. This paper identifies the worst-case rainfall scenarios that cause the most severe flooding loss through robust optimization, which constructs a new type of uncertainty set that builds upon IDF curves. Our analysis reveals that not all rainfalls on the same IDF curve are equal in terms of the incurred flooding loss. If the planners fail to account for the city's intrinsic infrastructure capabilities, they may pick a wrong 50-year rainfall to target and end up barely surviving a 2-year rainfall. Hence, we construct cost-indifferent isocurves that classify rainfall events according to the severity of the flooding loss, providing city planners with interpretable flooding risk mitigation guidelines. We demonstrate that cities should be savvy in achieving a balance between investments in stormwater retention flow rate and investments in stormwater retention capacity, as reflected in the composition of green and grey infrastructures. Under a budget constraint, green infrastructures should give way to grey ones, exchanging flow rates for a larger retention capacity to withstand severe rainfalls that have a longer duration and larger volume. Finally, we show that flooding loss ramps up quickly as the climate change risks are exacerbated, underscoring the urgency of taking the proposed integrated measures.

4 - Optimal Management of Renewable Energy Certificates (REC): A Reinforcement Learning Approach

Michael Lim, Seoul National University, Seoul, Korea, Republic of, Daeho Kim, Dong Gu Choi

Renewable Portfolio Standards (RPS), which require electricity suppliers to purchase Renewable Energy Certificates (RECs) from renewable energy generators to meet the standard levels, have led to the emergence of a brokerage service for renewable generators. Our research focuses on optimizing REC management for brokerage service providers. We use a Markov Decision Process framework to formulate the problem and identify key structural properties of the optimal policy. We then develop a deep reinforcement learning (DRL) algorithm and verify its performance based on real market data. Finally, we employ eXplainable AI techniques to understand the logic behind our black-box DRL algorithm

5 - Estimating the Impact of Climate Change: An Empirical Analysis of Smart Thermostat Data

Michael Blair, Wilfrid Laurier University, Waterloo, ON, Canada, Saed Alizamir, Shouqiang Wang

Using a rich micro-level dataset, we empirically analyze smart thermostat data to understand the relationship between households' thermostat settings and their ambient environment. Within this data we observe two key decision levers that households use to manage their home heating and cooling. Combining a variety of methodological tools including Dynamic Linear Models, random effects, and Bayesian Statistics, we develop models for these behavior levers and study how households are impacted by varying weather conditions. A critical insight from these models is that households typically react to the weather in a comfort-seeking manner, leading to increased energy usage. We also find that this behavior is more prevalent in households that routinely override the settings they have pre-programmed into the thermostat. With these insights in mind, we leverage well-established models to create realistic estimates of future weather conditions under a variety of climate change scenarios and combine these scenarios with our statistical models to estimate the impact of climate change on how households use their thermostats. Specifically, we predict future household behavior and investigate the resulting changes in energy usage for heating and cooling. These changes have a profound impact on the operations of the energy grid specifically relating to operating costs, grid capacity, and reliance on intermittent renewable energy sources. The insights derived from our research provide a valuable framework to evaluate and design interventions such as smart nudging or demand response management programs.

TE76

Regency - 707

Power Systems Operations and Planning V: Recent Advances

Invited Session

ENRE: Other Energy

Chair: Mengyi Sha, Tsinghua University, Beijing, N/A

Co-Chair: Wei Qi, Tsinghua University, Beijing, N/A, China, People's Republic of

1 - First-in-first-out (FIFO) or highest-energy-first-out (HEFO)? A tale of two electric vehicle (EV) battery swapping policies

Mengyi Sha, Tsinghua University, Beijing, China, People's Republic of, Wei Qi, Yuli Zhang

Battery swapping is gaining traction in cities and spurs wider EV adoption, as the business mode can alleviate range anxiety efficiently. In this paper, considering random service requests and random status of incoming batteries, we focus on the operations management of an individual battery swapping station, and propose mean-variance distributionally robust optimization (DRO) models to compare FIFO and HEFO swapping policies analytically. The results demonstrate that HEFO policy is superior to FIFO policy in most scenarios, except when all EVs wait for the recharged batteries to reach a target state-of-charge (SOC). We also verify the conclusions via numerical experiments using real-world data sets.

2 - Robust Grid-Vehicle Integration: A Machine Learning-Driven Optimization Approach

Ziliang Jin, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Jianqiang Cheng, Kai Pan, Zuo-Jun Shen, Yulan Wang

We examine a grid-vehicle integration system that uses vehicle-to-grid (V2G) technology to connect the grid and the electrical vehicle sharing system. Our focus is on the operation of this integration system under uncertainties. We formulate a two-stage robust mixed-integer program. We propose a machine learning-driven optimization approach that incorporates alternating direction method of multipliers (ADMM), machine learning, and strong valid inequalities. This approach significantly outperforms a commercial solver in both computational time and solution quality based on real data. Our findings suggest that in addition to enhancing the operation efficiency of the grid, V2G also promotes sustainability by reducing carbon emissions. However, the extent of this contribution varies under different power load patterns.

3 - Virtual Trading in Multi-Settlement Electricity Markets

Bo Yang, Columbia University, New York, NY, United States, Agostino Capponi, Garud Iyengar, Daniel Bienstock

In the electricity market, suppliers and load-serving entities (LSEs) enter into contracts to secure energy in the Day-Ahead (DA) market and later adjust for any discrepancy between contracted quantities and actual deliveries in the Real-Time (RT) market. This two settlement structure offers hedging benefits for both suppliers and LSEs but may lead to market inefficiencies. We develop a supply function equilibrium model to explore how virtual trading — a financial strategy that allows participants to speculate on price differences between the DA and RT markets without physical transactions — can mitigate these inefficiencies. Our analysis reveals that without virtual trading, LSEs often bid below their actual demand in the DA market, leading to DA prices that are lower than the expected RT prices. We show that introducing virtual trading narrows this price disparity and, with an increasing number of virtual traders, potentially eliminates it. Nonetheless, this mechanism prompts LSEs to further reduce their bids in the DA market, deviating from their true demand forecasts. Our analysis indicates that the presence of renewable energy suppliers leads LSEs to submit higher bids in the DA market, compared to scenarios with traditional energy suppliers only. We provide empirical support to our main model implications using data from the California and New York Independent System Operators.

TE77

Regency - 708

Learning and Optimization Techniques for Uncertain Systems

Invited Session

Computing Society

Chair: Rohit Kannan, Virginia Tech, Blacksburg, VA, United States

1 - A Generalizable Learning Approach to Accelerate Global Optimization of Quadratically-Constrained Quadratic Programs

Harsha Nagarajan, Los Alamos National Laboratory, Los Alamos, NM, United States, Erin George, Rohit Kannan, Deepjyoti Deka

Quadratically-constrained quadratic programs (QCQPs) are a broad class of non-convex optimization problems which occur in applications such as energy infrastructure networks and the pooling problem, but are often slow to solve in practice. Leveraging efficient MIP relaxations, and the sensitivity of partitioning choices, we propose the novel problem of "strong partitioning" to optimally partition variable domains "without" sacrificing global optimality guarantees. We propose to replace this expensive "ideal policy" with a machine learning (ML) approximation for homogeneous families of non-convex QCQPs. Further, we construct a novel end-to-end loss function so that the desired attributes of an output partitioning scheme can be learned directly, without reference to a pre-selected policy. Additionally, we introduce a graph-based method to produce a single ML model which works on a variety of different sizes. Numerical results demonstrate the ML model's ability to learn partitioning policies which generalize well, even to unseen problem sizes, significantly reducing the run times of partitioning-based iterative algorithms.

2 - Learning to Generate Cutting Planes with Miplearn

Alinson Santos Xavier, Argonne National Laboratory, Lemont, IL, United States

While many classes of cutting planes described in the literature have been effective in terms of gap closure, they are often computationally demanding to generate and offer only modest improvements in running times. Taking into account that many discrete optimization problems are solved repeatedly with only slight variations in input data, in this talk we investigate the usage of machine learning (ML) methods to accelerate the selection and generation of such cuts. We start by introducing MIPLearn, an extensible open-source framework which uses machine learning (ML) to enhance the performance of state-of-the-art MIP solvers. The framework is compatible with multiple MIP solvers (e.g. Gurobi, CPLEX, SCIP, HiGHS), multiple modeling languages (JuMP, Pyomo, gurobipy) and supports user-provided ML models. We then present computational experiments on using ML to accelerate the generation of rank-1 Gomory mixed-integer (GMI) cuts from multiple tableau bases.

3 - Power Systems Optimization Under Uncertainty

Rahul Gupta, Georgia Institute of Technology, Atlanta, GA, United States

The uncertainties within the power system's operation and planning can be broadly classified into two categories. The first refers to cases where uncertainty lies in power injections, such as demand or generation, which cause stochastic fluctuations in the power system's state, such as nodal voltages. The second kind refers to cases where uncertainty exists in the power system models themselves, for example, inaccurate network parameters such as line impedance or topology. Such uncertainties lead to inaccuracies in estimating or predicting power system states. Overall, both kinds of uncertainty lead to challenges in power system operation. These challenges are particularly linked to the regulation of network states, such as nodal voltages that must be kept within statutory limits, as well as managing power flows along lines and transformers within rated capacities. Such challenges prompts the important question of how to make optimal decisions amidst uncertainty. This talk will present different methods to tackle the above two kinds of uncertainties using stochastic and robust optimization algorithms. Additionally, the talk will also present two different power system applications for addressing these uncertainties.

4 - A Computational Study of Cutting-plane Methods for Multi-stage Stochastic Integer Programs

Akul Bansal, Northwestern University, Evanston, IL, United States, Simge Kucukyavuz

We report a computational study of cutting plane algorithms for multi-stage stochastic mixed-integer programming models with the following cuts: (i) Benders', (ii) Integer L-shaped, and (iii) Lagrangian cuts. We first show that Integer L-shaped cuts correspond to one of the optimal solutions of the Lagrangian dual problem, and, therefore, belong to the class of Lagrangian cuts. To efficiently generate these cuts, we present an enhancement strategy to reduce time-consuming exact evaluations of integer subproblems by alternating between cuts derived from the relaxed and exact computation. Exact evaluations are only employed when Benders' cut from the relaxation fails to cut off the incumbent solution. Our preliminary computational results show the merit of this approach on multiple classes of real-world problems.

TE78

Regency - 709

Federated Learning and Optimization: III

Invited Session

Computing Society

Chair: Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

1 - Exploring Fairness, Efficiency, and Incentives in Federated Learning through Game Theory and Social Choice

Aniket Murhekar, University of Illinois, Urbana-Champaign, Urbana, IL, United States

Federated learning (FL) has emerged as a powerful scheme to facilitate the collaborative learning of models amongst agents holding their own private data. While agents benefit from sharing data through as it results in accurate models, they also incur costs associated with data sharing such as communication or privacy costs. This can lead to agents being strategic about their contribution to achieve an optimal tradeoff between their learning payoff and data sharing cost, and can lead to free-riding.

To study such issues of incentives arising in FL, we consider the existence of Nash equilibrium (NE) in a simple collaborative learning model. We show the existence of Nash equilibrium (NE) under mild assumptions on agents' payoff and costs. However, some of the NE may be bad in terms of overall welfare for the agents, implying little incentive for some fraction of the agents to participate in the learning. To remedy this, we design a budget-balanced mechanism involving payments to the agents, that ensures that any p-mean welfare function of the agents' utilities is maximized at NE. In addition, we introduce a FL protocol FedBR-BG that incorporates our budget-balanced mechanism, utilizing best response dynamics. Our empirical validation on MNIST and CIFAR-10 substantiates our theoretical analysis. We show that FedBR-BG outperforms the basic best-response-based protocol without additional incentivization, the standard federated learning protocol FedAvg, as well as a recent baseline MWFed in terms of achieving superior p-mean welfare.

2 - Distributed Randomized Zeroth-Order Gradient Tracking Methods for Stochastic MPECs

Mohammadjavad Ebrahimi, Rutgers University, Piscataway, NJ, United States, Uday Shanbhag, Farzad Yousefian

We consider a class of hierarchical multi-agent optimization problems over networks where agents seek to compute an approximate solution to a single-stage stochastic mathematical program with equilibrium constraints (MPEC). MPECs subsume several important problem classes including Stackelberg games, bilevel programs, and traffic equilibrium problems, to name a few. Our goal in this work is to provably resolve stochastic MPECs in distributed regimes where the agents only have access to their local objectives and an inexact best response to the lower-level equilibrium problem. To this end, we devise a new method called randomized smoothed distributed zeroth-order gradient tracking (rs-DZGT). This is a novel gradient tracking scheme where agents employ a zeroth-order implicit scheme to approximate their (unavailable) local gradients. Leveraging the properties of a randomized smoothing technique, we establish the convergence of the method and derive complexity guarantees for computing a stationary point of an optimization problem with a smoothed implicit global objective. We also provide preliminary numerical experiments where we compare the performance of rs-DZGT on networks under different settings with that of its centralized counterpart.

3 - Federated Convex Bilevel Optimization: A Universal Regularized Scheme with Guarantees

Farzad Yousefian, Rutgers University, Piscataway, NJ, United States, Mohammadjavad Ebrahimi, Yuyang Qiu, Shisheng Cui

We study a bilevel federated learning (FL) problem, where clients cooperatively seek to find among multiple optimal solutions of a primary distributed learning problem, a solution that minimizes a secondary distributed global loss function. This problem has attracted increasing attention in machine learning, in particular, in over-parameterized learning and hyperparameter optimization. Despite some recent progress, communication-efficient FL methods equipped with complexity guarantees for resolving this problem are primarily absent. Motivated by this lacuna, we propose a universal regularized scheme and derive promising error bounds in terms of both the lower-level and upper-level loss functions. Leveraging this unifying theory, we then enable existing FL methods, including FedAvg and SCAFFOLD, to solve the corresponding bilevel FL problem, and derive novel communication complexity guarantees for each method. Intriguingly, the universal scheme can be employed to provably enable many other state-of-the-art optimization methods to address the bilevel problem. We present numerical experiments for over-parameterized convex neural networks on MNIST and CIFAR-10 datasets.

4 - Scalable Privacy-Preserving Federated Learning for Electric Load Forecasting

Shourya Bose, University of California Santa Cruz, Santa Cruz, CA, United States, Yu Zhang, Kibaek Kim

Federated learning (FL) has emerged as a powerful approach for training models using data distributed across numerous devices. In this study, we address the application of privacy-preserving federated learning (PPFL) to train electric load forecasting models across a vast network of smart meters, with enhanced privacy assurances compared to conventional FL methods. Our approach leverages differential privacy and personalized layers (PLs) to effectively manage data heterogeneity while safeguarding against potential data leakage from model weights.

Tuesday, October 22, 5:25 PM - 6:15 PM

Summit - Ballroom 2

Edelman Reprise: Molslinjen

Plenary/Keynote Session

Keynote

Chair: Rajesh Tyagi, GE Global Research, Plainsboro, NJ, United States

1 - Molslinjen – Passenger Ferry Operations in the Digital Era: Forecasting & Revenue Management at Molslinjen

Pierre Pinson, Halfspace / Imperial College London, Copenhagen, Denmark

Halfspace and Molslinjen have partnered to develop and operate a successful forecasting and revenue management toolbox for data-driven operation of ferries in Denmark, rolled out operationally since 2020. This has resulted in $2.6-3.2M yearly savings and been a significant contributor to a 3% reduction in fuel costs and emissions. This toolbox relies on the latest advances in machine learning for forecasting and ORMS approaches to revenue management. The potential for generalizing to the global ferry industry is significant, with an impact on both revenues and ESG criteria.

Tues Keynote

Summit - Ballroom 1

OR/AI Applications in the (Bio)Pharmaceutical Industry: A Vision and Future Research Directions

Plenary/Keynote Session

Keynote

Chair: Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

1 - OR/AI Applications in the (Bio)Pharmaceutical Industry: A Vision and Future Research Directions

Tugce Martagan, Eindhoven University of Technology, Eindhoven, Netherlands

The application of Operations Research and Artificial Intelligence (OR/AI) methodologies has successfully helped transform several industries. However, there is still a significant need for more OR/AI research in the (bio)pharmaceutical industry. Until recently, competitive advantage in the (bio)pharmaceutical industry has been driven primarily by advances in life sciences. However, with increasing demand and competition, the industry is becoming more sensitive to cost efficiency, supply chain resilience, and equitable patient access to medicines.

In this presentation, we will reflect on how OR/AI applications can help advance the (bio)pharmaceutical industry. We will begin with an overview of current industry trends, needs and challenges. We will also discuss some examples from manufacturing and supply chain applications, based on our work presented in the 2022 Franz Edelman competition. Finally, we will propose a research roadmap and a portfolio of future research directions for OR/AI applications in (bio)pharmaceutical manufacturing.

Wednesday, October 23, 8:00 AM - 9:15 AM

WA01

Summit - 320

AI-Human Interation

Invited Session

Service Science

Chair: Kejia Hu, University of Oxford Said Business School, Oxford, United Kingdom

Co-Chair: Yuhan Su, Tianjin University, Tianjin, 300072

1 - How Matching Algorithm Innovations Impact Platform Economy Ecosystem: Empirical Evidence from Ridesharing Platforms

Mengwei Qu, University of Connecticut, Storrs, CT, United States, Bowen Lou, Chen Liang

Matching algorithms serve as the backbone of ridesharing platforms, facilitating efficient and timely connections between passengers and drivers. These sophisticated algorithms meticulously optimize various factors, including geolocation, supply and demand dynamics, driver allocation strategies, and routing and traffic conditions, to facilitate successful matches. However, the pursuit of optimization often entails imposing extensive algorithmic control, potentially leading to inefficiencies in matching processes. Our research delves into the realm of matching algorithm innovation, particularly through the development of a reduced control approach that operates at an aggregate level rather than micromanaging individual parameters. Leveraging a Triple Difference-in-Differences model, we evaluate the effects of these innovations. Our findings reveal a notable decrease in riders' wait times, signifying a tangible enhancement in service efficiency. Additionally, we uncover significant improvements in drivers' earnings, earnings efficiency, and overall rider satisfaction metrics. Furthermore, innovative matching algorithms also influence drivers' pick-up location decisions, suggesting a positive impact on drivers' preference for the matching algorithm. Our study provides valuable empirical evidence elucidating the profound impact of matching algorithm innovations on the ridesharing platform ecosystem. Beyond insights, it furnishes actionable managerial recommendations for online platforms, highlighting the efficacy of tailored matching algorithms in elevating service quality, fostering efficiency, and potentially reducing operational costs. These findings pave the way for future advancements in ridesharing technology and algorithmic optimization strategies.

2 - Strategic Value of Firm-Specific Knowledge in Big Data Analytics: Insights from S&P 500 Firms

Yuhan Su, Tianjin University, Tianjin, China, People's Republic of, Yuanyang Liu, Kejia Hu, Xiande Zhao

In the era of big data, firms increasingly value Big Data Analytics (BDA) for its potential to enhance efficiency and drive superior performance. While previous research has emphasized the value of Big Data Analytics Capabilities (BDAC), this study addresses the gap in understanding the impact of firm-specific knowledge within BDA human capital. We employ a comprehensive dataset for S&P 500 firms, comprising workforce data on data scientists, financial data, and job advertisement data from 2011 to 2019, to analyze firm-specific knowledge, inflow, and outflow of data scientists alongside firm performance metrics.

Our findings indicate that data scientists' firm-specific knowledge significantly improves Return on Assets (ROA), whereas IT workers' firm-specific knowledge has a negligible effect. Additionally, we categorize companies based on their level of data analytics and integration capabilities (High, Centralized Analytical, Decentralized Analytical, Low) and identify that the positive impact of data scientists' firm-specific knowledge and inflow is maximized during the Centralized Analytical stage of data integration strategies. This suggests that firms can enhance the benefits derived from data scientists' tenure and inflow by focusing on their business integration capabilities. Our study provides empirical evidence supporting the strategic value of firm-specific human capital in the context of BDA, offering insights for firms to optimize their data integration strategies and achieve a competitive advantage.

WA02

Summit - 321

Complex System Modeling, Monitoring, and Decision-making

Invited Session

Quality, Statistics and Reliability

Chair: Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Xiaochen Xian, University of Florida, Gainesville, FL, United States

1 - A Bayesian Jump Model-Based Pathwise Sampling Approach for Online Anomaly Detection

Dongmin Li, University of Florida, Gainesville, FL, United States, Miao Bai, Di Wang, Xiaochen Xian

Moving vehicle-based sensors (MVSs) have received growing attention for real-time anomaly detection in various applications such as wildfire and oil spill detection. To tackle challenges due to the spatial covariance structure among observations, uncertainties under partial observations, as well as the physical MVS movements, we propose a Bayesian jump model-based pathwise sampling approach to detect abrupt changes in an area of interest in real time using MVSs. Specifically, we integrate a jump-model based Bayesian scheme, the upper confidence bound algorithm, and mathematical optimization in a unified manner to exploit spatial correlation and real-time partial observations, to handle uncertainties of noisy observations, limited observability, and anomaly occurrences, and to adaptively coordinate the routes of multiple MVSs for quick anomaly detection. We perform theoretical investigations and conduct simulations to confirm the exceptional effectiveness of the method we propose. A case study for early wildfire detection demonstrates that our proposed method outperforms benchmark methods, which contributes to the reduction of the area of affected land and wildfire-related costs.

2 - Reinforced Scan: a Reinforcement Learning Enabled Scanning Strategy in Pbf Additive Manufacturing

Chaoran Dou, Virginia Tech, Blacksburg, VA, United States, Jihoon Chung, Raghav Gnanasambandan, Yuhao Wu, Zhenyu Kong

Additive Manufacturing (AM) is an innovative technology that fabricates parts layer by layer. However, in powder bed fusion (PBF), printed metal parts often exhibit residual stresses, deformations, and other defects due to uneven temperature distribution during the printing process. To address this, an optimized scanning sequence within each layer can help mitigate temperature inconsistencies. Traditional optimization methods are based on domain knowledge, employing try-and-error or heuristic methods. Nonetheless, these methods are not universal and cannot achieve the optimal. The challenge of improving the scanning strategy is the large search space for optimizing the scanning sequence for the scanning tracks within the layer. To fill this gap, this work proposes an innovative scan strategy aiming to optimize scanning sequences for achieving uniform temperature distribution in PBF. The proposed Reinforced Scan approach uses reinforcement learning methods to determine the scanning sequence with a customized reward function intelligently. This reward function not only considers the temperature variance but also the spatial uniformity of temperature distribution. This method can significantly reduce the computational burden involved in scanning sequence optimization. The effectiveness of the proposed Reinforced Scan is validated through Netfabb Local Simulation involving laser scanning on a Ti64 thin plate, where its performance is compared with existing heuristic scan sequences. The simulation results demonstrate that Reinforced Scan yields superior outcomes, achieving reduced residual stress compared to conventional heuristic methods.

3 - Boundary Quality Characterization and Modeling for Printed 2D Products Using Printing Primitives

Minghao Gu, University of Southern California, Los Angeles, CA, United States

Surface quality characterization is essential for design, manufacturing and product functionality. Although a wealth of literature has been developed for surface quality characterization in mass production, it faces unique challenges in additive manufacturing (AM) due to one-off fabrication of a large variety of products with complex geometries. Surface quality of AM products side surfaces can be viewed as the stacking-up and combination of boundary quality for each fabricated layer. Boundary quality characteristics can vary with built geometries and covariates such as size, built location and orientation, which makes existing methods not adaptable to frequent design changes in AM. To address these challenges, we proposed a novel boundary surface quality characterization and modeling approach based on printing primitives. We partition a 2D design into a finite amount of printing primitives. The in-plane boundary quality patterns on each type of primitives are characterized through tensor-product modeling with integration of domain knowledge to reflect the impact of individual covariates. Simulation studies demonstrate the capability of the proposed model for learning the stochastic process governing the boundary quality distribution. We also illustrate the application of the proposed model for analyzing real printed samples produced by an AM process.

4 - Conformal Spatiotemporal Forecasting and Application in Energy Load Prediction

Shiyuan Piao, The Hongkong University of Science and Technology (Guangzhou), Guangzhou, China, People's Republic of, Fugee Tsung

Conformal prediction is a statistical framework that provides valid probabilistic predictions with well-calibrated confidence sets. This approach is particularly useful in spatial-temporal forecasting, where the goal is to predict future values of a variable based on its past values and spatial relationships with other variables. In the context of energy load forecasting, conformal prediction can offer significant advantages over traditional statistical and machine learning methods.

Spatial-temporal data in energy load forecasting is characterized by complex patterns and dependencies that evolve over time and space. Conventional forecasting methods often struggle to capture these intricate dynamics, leading to suboptimal predictions. However, conformal prediction, with its non-parametric nature, can adaptively learn from the data without making strong assumptions about the underlying distribution. This flexibility allows conformal methods to better model the spatial-temporal dependencies and provide more accurate and reliable forecasts.

The use of conformal prediction in energy load forecasting has several benefits. It provides a principled way to quantify uncertainty, which is crucial for risk management and decision-making in the energy sector. Additionally, conformal methods can handle non-stationarity and changing patterns in energy consumption, making them robust to shifts in demand that might be caused by economic, social, or technological factors.

In conclusion, conformal prediction offers a powerful and flexible approach to spatial-temporal forecasting, with promising applications in energy load forecasting. By leveraging the inherent structure of spatial-temporal data, conformal methods can deliver more accurate and reliable predictions, supporting the efficient and sustainable management of energy resources.

WA03

Summit - 322

Theoretical Perspectives and Statistical Approaches to Robust Inference

Invited Session

Quality, Statistics and Reliability

Chair: Liyan Xie, University of Minnesota, Shenzhen

Co-Chair: Taposh Banerjee, University of Pittsburgh, Pittsburgh, PA, United States

1 - Distributionally Robust Multi-Hypothesis Testing with Moment Constrained Uncertainty Sets

Akshayaa Magesh, UIUC, Urbana, IL, United States, Venugopal Veeravalli, Zhongchang Sun, Shaofeng Zou

The problem of robust multi-hypothesis testing in the Bayesian setting is studied in this paper. Under the m hypotheses, the data-generating distributions are assumed to belong to uncertainty sets constructed through some moment functions, i.e., the sets contain distributions whose moments are centered around empirical moments obtained from some training data sequences. The goal is to design a test that performs well under all distributions in the uncertainty sets, i.e., a test that minimizes the worst-case probability of error over the uncertainty sets. Insights on the need for optimization-based approaches to solve the robust testing problem with moment constrained uncertainty sets are provided. The optimal (robust) test based on the optimization approach is derived for the case where the observations belong to a finite-alphabet. When the size of the alphabet is infinite, the optimization problem is infinite-dimensional and intractable, and therefore a tractable finite-dimensional approximation is proposed, whose optimal value converges to the optimal value of the original problem as the size of the dimension of the approximation goes to infinity. A robust test is constructed from the solution to the approximate problem, and guarantees on its worst-case error probability over the uncertainty sets are provided. Numerical results are provided to demonstrate the performance of the proposed robust test.

2 - Non-Asymptotic Convergence of Discrete-TIME Diffusion Models: New Approach and Improved Rate

Yuchen Liang, The Ohio State University, Columbus, OH, United States, Peizhong Ju, Yingbin Liang, Ness Shroff

The denoising diffusion model has recently emerged as a powerful generative technique that converts noise into data. While theoretical convergence guarantee has been extensively studied by discretizing some continuous-time diffusion processes, many generative samplers in real applications directly employ a discrete-time diffusion process. However, existing techniques for analyzing discretized continuous-time processes are not applicable to discrete-time processes, and for such discrete-time processes a convergence result has only been obtained for distributions with bounded support. In this paper, we establish the convergence guarantee for substantially larger classes of distributions under such discrete-time diffusion processes and further improve the convergence rate for distributions with bounded support. In particular, we first establish the convergence rates for both smooth and general (possibly non-smooth) distributions having finite second moment. We then specialize our results to a number of interesting classes of distributions with explicit parameter dependencies, including distributions with Lipschitz scores, Gaussian mixture distributions, and any distributions with Gaussian perturbation. We further propose a novel accelerated sampler and show that it improves the convergence rates of the corresponding regular sampler by orders of magnitude with respect to all system parameters. For distributions with bounded support, our result improves the dimensional dependence of the previous convergence rate by orders of magnitude. Our study features a novel analytical technique that constructs a tilting factor representation of the convergence error and exploits Tweedie's formula for handling Taylor expansion power terms.

3 - Revisit the Trade-off Between Accuracy and Robustness in Adversarial Training

Qunzhi Xu, Georgia Institute of Technology, Atlanta, GA, United States

The ability of generalizing to unseen, adversarially perturbed data is an important metric for evaluating the effectiveness of methods in improving the robustness of machine learning models. In this paper, we study the generalization properties of adversarial training (AT) and adversarial regularization (AR) in linear regression. Specifically, we consider a standard linear regression model and a perturbed linear regression model whose coefficient might change. By studying the asymptotic performance of AR and AT in these two models, we show that AR can generalize better in the sense of achieving a lower adversarial risk as compared to AT. Extensive simulation studies and real-world data analysis on different machine learning models are conducted to validate our theoretical results.

4 - Score-based approach to robust quickest change detection

Taposh Banerjee, University of Pittsburgh, Pittsburgh, PA, United States, Sean Moushegian, Vahid Tarokh

Score-based algorithms are proposed for the quickest detection of changes in unnormalized statistical models. These are models where the densities are known within a normalizing constant. These algorithms can also be applied to score-based models where the score, i.e., the gradient of log density, is known to the decision-maker. Performance analysis is provided for these algorithms and compared with their classical counterparts. It is shown that strong performance guarantees can be provided for these score-based algorithms where the Kullback-Leibeler divergence between pre- and post-change densities is replaced by their Fisher divergence. Robust solutions are also obtained and analyzed.

WA04

Summit - 323

Innovative Optimization Techniques and Applications

Contributed Session

Chair: Iris Forma, Tel-Aviv Afeka Academic College of Engineering, School of Industrial Engineering and Management, Tel Aviv - Yafo, Israel

1 - Hybrid classical-quantum approaches based on the column-generation framework for hard combinatorial problems: Examples and Benchmarks

Victor Drouin-Touchette, Université de Sherbrooke, Sherbrooke, QC, Canada

In this talk I will present an overview of various hybrid classical-quantum algorithms that are inspired by classical column-generation procedures developed in the field of operations research. In addition to leveraging fast classical solvers of linear programs, we use a quantum sampler based on neutral-atom platforms to heuristically propose new columns to add to a given combinatorial optimization problem. Benchmarks on the minimum vertex coloring problem and the traveling salesman problem will be presented, which together show that these hybrid algorithms can deliver good solutions in few iterations. I will also present a comparison with different state-of-the-art classical and quantum approaches and comment on the potential benefits of a quantum-enhanced solution workflow.

2 - Heuristic Emphasis in FICO Xpress

Leona Gottwald, FICO, Berlin, Germany

This talk will dive into primal heuristics for mixed-integer models in the FICO Xpress Solver.

We will highlight recent advances, new MIP heuristics, and take a closer look on the heuristic emphasis mode of Xpress. The heuristic emphasis mode aims at finding feasible solutions early during the search.

3 - Complexity-Optimal and Parameter-Free First-Order Methods for Finding Stationary Points of Composite Optimization Problems

Weiwei Kong, Google, New York, NY, United States

This talk presents a "parameter-free" and "complexity-optimal" accelerated proximal descent method for finding stationary points of nonconvex composite optimization problems. The main advantages of this method are that (i) it does not require knowledge of the global topological properties of its objective function and (ii) it obtains optimal iteration complexity bounds for nonconvex problems and nearly optimal bounds (up to logarithmic terms) for convex problems, without any parameter tuning. To the author's knowledge, this is the first method in the literature to obtain (i) and (ii) simultaneously. Another point of interest is the new set of inexact proximal termination conditions used in developing the method's outermost iterations.

This talk also describes how the proposed method can be leveraged in other optimization frameworks, such as min-max smoothing and penalty frameworks for constrained programming, to create more specialized parameter-free methods. Finally, we present several numerical experiments to support the method's practical viability.

4 - Application of federated optimization on generating stochastic scenarios

Mehrdad Mohammadi, Auburn University, Auburn, AL, United States, Alexander Vinel

This presentation introduces a new stochastic optimization approach using federated learning. This decentralized machine learning approach enables model training across multiple devices or locations while keeping data local, which is being utilized to generate different scenarios based on predicted values.

Data is distributed between clients, and the federated learning approach will optimize each client's model and aggregate them into the main model.

This approach leverages the power of federated learning, enabling optimization techniques tailored for stochastic scenario generation. This method optimizes decision-making processes in complex, dynamic environments by seamlessly integrating distributed data sources.

5 - Optimize Shading Planning in Cities

Iris Forma, Tel-Aviv Afeka Academic College of Engineering, School of Industrial Engineering and Management, Tel-Aviv, Israel

This research introduces an optimization approach for planning shading in cities, addressing the significant challenges posed by heat waves and Urban Heat Islands in the coming decades. Recognizing the imperative need to cool cities, there is a growing awareness of the need for solutions. While reforestation stands out as an efficient strategy due to the unique benefits of trees for the environment and human health, it may not be feasible in all urban areas. Consequently, providing shade becomes crucial even in areas where tree planting is impractical. Decision-makers in such locations can deploy artificial shadings tailored to urban environments, such as shade sails and pergolas.

This research presents an innovative optimization model that integrates both natural and artificial shading. It considers climatic, geographic, economic, and demographic parameters. The primary objective is to maximize climatic comfort by optimizing decisions regarding the placement and surface area of both natural and artificial shading in open and urban spaces.

Comparing our mathematical model with conventional decision-making in 10 scenarios, real data from the Mapping and Geospatial Information Unit of the Tel Aviv municipality revealed a 33% improvement in shaded areas and a 68% increase in the number of shaded locations, on average. Additionally, we provide a budget sensitivity analysis and discuss ongoing research.

WA05

Summit - 324

Resource Management in Services

Contributed Session

Chair: Cong Wang, Peking University, Beijing, N/A, China, People's Republic of

1 - Balancing Privacy Preservation and Data Value in Smart Meters: A Correlation-Aware Differential Privacy Approach

Cong Wang, Peking University, Beijing, China, People's Republic of

The wide-scale deployment of smart grids has revolutionized data-driven power management, but the collection of fine-grained and correlated electricity consumption data recorded by smart meters poses a potential threat to user privacy. To address this, we propose a Correlation-Aware Differential Privacy (CaDP) method to balance user privacy with data utility in smart meter data. CaDP addresses privacy concerns by considering data correlation, which allows for more efficient use of the privacy budget. It presents a bi-objective optimization to balance privacy and data value, providing a practical framework for selecting privacy parameters. Real-world data evaluation shows CaDP's effectiveness in preserving privacy while maintaining data utility for tasks like load forecasting, outperforming other methods. The work contributes to the theory and practice of cybersecurity by offering both theoretical insights and practical guidelines for managing the privacy-data value tradeoff in smart grids.

2 - Using Bio-Inspired Reconfigurations to Enhance the Resilience of Electric Power Systems: A Decentralized Approach

Tianye Wang, George Washington University, Washington, DC, United States

This study introduces a robust decentralized control algorithm designed to significantly enhance the resilience of electrical power systems. The innovation integrates key characteristics of biological cell development, notably environmental awareness, decentralization, and redundancy. By mimicking the adaptive and autonomous features of cellular systems, the algorithm empowers a networked microgrid, to autonomously detect, respond to, and recover from localized disruptions without centralized oversight. To validate its efficacy, a stochastic cost-minimization model is employed to demonstrate improved resilience metrics, such as shorter recovery time and lower economic loss, under a range of failure scenarios, including extreme weather events. Preliminary results indicate a significant enhancement in the system's ability to adapt to and recover from adverse conditions, with potential implications for the design of future resilient power infrastructure. The algorithm's scalability and adaptability to different grid configurations are also discussed, highlighting its broad applicability and potential for real-world implementation.

3 - Regularized Benders Decomposition for High Performance Capacity Expansion Models

Filippo Pecci, Princeton University, Princeton, NJ, United States, Jesse Jenkins

We consider electricity capacity expansion models, which optimize investment and retirement decisions by minimizing both investment and operation costs. In order to provide credible support for planning and policy decisions, these models need to include detailed operations and time-coupling constraints, and allow modeling of discrete planning decisions. Such requirements result in large-scale mixed integer optimization problems that are intractable with off-the-shelf solvers. Hence, practical solution approaches often rely on carefully designed abstraction techniques to find the best compromise between reduced temporal and spatial resolutions and model accuracy. Benders decomposition methods offer scalable approaches to leverage distributed computing resources and enable models with both high resolution and computational performance. Unfortunately, such algorithms are known to suffer from instabilities, resulting in oscillations between extreme planning decisions that slows convergence. In this study, we implement and evaluate several level-set regularization schemes to avoid the selection of extreme planning decisions. Using a large capacity expansion model of the Continental United States with over 70 million variables as a case study, we find that a regularization scheme that selects planning decisions in the interior of the feasible set shows superior performance compared to previously published methods, enabling high-resolution, mixed-integer planning problems with unprecedented computational performance.

4 - Empowering Privacy Policy Understanding and Accessibility in Smart Home Devices

HILAL PATACI, University of Texas at San Antonio, San Antonio, TX, United States, Ke Yang, Anthony Rios

Despite nearly 70 million U.S. households using smart home technology, significant gaps in the clarity and accessibility of privacy policies hinder informed data use decisions. Alarmingly, recent studies indicate that almost 50% of devices in the U.S. market lack a privacy policy. This lack of transparency, combined with the complexity of existing policies, poses serious privacy and security concerns. Our research seeks to address these issues by enhancing the interpretability of these legal documents, ensuring they are accessible and understandable to a diverse range of users, including vulnerable populations.

We introduce a novel QA system specifically tailored for privacy policies associated with smart home technologies. The research is structured around three main objectives: enhancing data collection by creating a comprehensive smart-home device-specific QA dataset; evaluating the performance of existing QA systems; and developing robust, inclusive QA models that address the nuances of privacy policy queries. The novel QA system includes the integration of privacy policies with user-generated questions and vulnerability information from NIST's National Vulnerability Database, employing novel data curation techniques to enrich the QA dataset. By leveraging state-of-the-art NLP technologies and innovative algorithmic frameworks, we aim to improve the accuracy and fairness of QA systems, making privacy information more accessible and understandable for diverse users.

The broader impact of this work extends beyond technological advancements, fostering trust and compliance with privacy regulations among users. By empowering individuals to make informed decisions about their data, this research contributes to the fields of data privacy, algorithmic fairness, and smart home technology.

5 - The Value of Flexible Resources

shilei niu, Department of Economics, Faculty of Arts, University of Waterloo, Waterloo, ON, Canada, Tony Wirjanto

Flexible resources are critical for the power system reliability due to the high penetrations of variable renewable energy (VRE) during the decarbonization of the electricity industry. Using a real options approach, this paper develops a general model to value a class of assets with the power generation, storage and ramp capability. Within the non-market valuation framework, we adopt the no-arbitrage principle to measure the marginal value of flexible ramping through the shadow price which is the fair value implied from the electricity markets. To evaluate the incentives to invest in flexible resources and provide flexible ramping products (FRPs) in the power markets, we conduct extensive elasticity and sensitivity analyses to understand how the scarcity and cost of resources for power production and the electricity price signals affect the asset value and shadow price. Our numerical results show that the plant value decreases, but the shadow price increases exponentially as the ramping constraints become more restrictive. Both the flexible ramping up (FRU) and flexible ramping down (FRD) are valued, while the FRU has much higher value to the power generator. The scarcity and cost of resources and the electricity price characteristics have significant impacts on the asset value and shadow price; however, the magnitudes and directions of the impacts vary among these parameters. This research will facilitate the investment decisions to the flexible resources, development on the FRPs market and implementation of ramping regulations designed to promote the integrity of the ecosystem and climate while meeting the power system reliability.

WA06

Summit - 325

Advances in Machine Learning Applications

Contributed Session

Chair: Keli Xiao, Stony Brook University, Stony Brook, NY, United States

1 - A Newly Designed Data Envelopment Analysis Approach that Ranks a Set of Mutual Funds and Mutual Fund Companies

Richard Muszynski III, Wilkes University, Wilkes-Barre, PA, United States

One of the many great investment choices for investors has been the plethora of mutual funds. The popularity of mutual funds arises from their diverse selection of stocks, bonds and other securities that reduces an investors overall risk of losing capital. Since there exists an abundance of mutual fund choices, the investors' goal of selecting mutual funds that match with their beliefs or expectations becomes an ever-daunting task. We provide analysis regarding which mutual funds an investor should invest in by using a nonparametric linear-programming approach called data development analysis (DEA). The goal of the present paper is to analyze and rank a set of mutual funds across several mutual fund companies by solving the input-oriented DEA model (Banker et al., 1984) and the output-oriented DEA model (Banker et al., 1984). After solving both models, we obtain distance measures and compute the magnitude distance for each country across multiple dimensions in order to calculate efficiency scores for each respective mutual fund. Rankings are determined by sorting the efficiency scores for each respective mutual fund. After ranking the set of mutual funds for each mutual fund company, we then rank mutual fund companies by taking averages, medians, and standard deviations of the efficiency scores of individual mutual funds of each company to determine overall and consistent mutual fund company performance. We also demonstrate why standardizing data is crucial in order to create equal weighted factors that may rank mutual funds.

2 - Does Size Matter? Short-Term Momentum versus Long-Term Reversal Returns in Investment Strategies

Nafise Aalipour, University of Rhode Island, Kingston, RI, United States, Georges Tsafack

Stock market direction is challenging to predict. However, it has been well documented that well-performing stocks tend to keep up in the short-term while some reversal is observed in the long run. We separately investigate those phenomena and their relation to the market size of stocks. Interestingly, our research indicates that long-term reversal is predominantly driven by small market cap stocks, whereas momentum is more commonly observed in medium and large size stocks. Moreover, our findings hold true whether we apply traditional measures of momentum and reversal or utilize risk-adjusted metrics such as the Sharpe ratio or alpha for the asset allocation process. Our analysis of data spanning from 1927 to 2022 reveals that these trends have become more pronounced post-1966, coinciding with advancements in financial research tools such as CAPM-alpha and the Sharpe ratio.

3 - Enhancing Long-Term Decision-Making for Social Welfare: A Multi-Agent Reinforcement Learning Approach

Keli Xiao, Stony Brook University, Stony Brook, NY, United States, PENGZHAN GUO, Jingyuan Yang

Sequential decision-making with a long-term perspective is important in various domains, but existing decision support systems often prioritize short-term, individual solutions. This research explores the potential of AI techniques, specifically reinforcement learning, to enhance long-term decision-making by maximizing social benefit. We address the challenge of multi-user scenarios with limited social resources by formulating a generalized problem and translating it into a multi-agent reinforcement learning framework. Our approach emulates a social decision system, balancing individual objectives with collective welfare. To manage the computational complexity of multi-user decision-making, we propose a deep learning-guided, performance-weighted search algorithm. This algorithm leverages deep neural networks and performance-based weightings to enhance computational efficiency and adaptability, dynamically adjusting the search strategy to different path requirements, such as talent mobility and mobile route recommendations. We provide evidence from multi-agent reinforcement learning simulations to demonstrate the effectiveness and efficiency of our method. The simulations, using real-world data in tasks like sequential route recommendation and long-term career planning, show that our approach outperforms state-of-the-art techniques. Robustness tests using synthetic data confirm the stability and reliability of our method.

WA10

Summit - 330

Advancements in Supply Chain Resilience and Performance

Contributed Session

Chair: Wenxin Zhang, syracuse university, 721 University Ave, syracuse, 13244

1 - Supply Chain Mapping with Synthetic Data Powered by Large Language Models

Jaewon Kim, Korea University, Seoul, Korea, Republic of, Eunbi Kim, Dongsu Kim, Yoojoong Kim, Taesu Cheong

Supply chain mapping is crucial for global companies to identify and assess potential risks in their supply chains. While previous research has explored the use of Natural Language Processing (NLP) techniques for automated supply chain mapping, these methods face challenges due to data scarcity and the high cost of data annotation. This study proposes a novel approach to enhance the accuracy and efficiency of supply chain mapping using synthetic data generated by large language models (LLMs). The proposed methodology involves a two-part process. First, a multi-turn prompt guides LLMs to generate relation triples of company-relation-company. The prompt includes a structured output template to ensure the generated text, including corresponding synthetic sentences, adheres to a predefined schema for automated parsing. Second, these synthetic sentences and their corresponding relation triples are used to fine-tune a BERT-based relationship classifier. The results show that the BERT classifier surpasses previous state-of-the-art results in NLP-based supply chain mapping, and an innovative method using special tokens at the position of company names further improves performance. Augmenting the training dataset with LLM-generated synthetic examples proves to be a viable approach, especially when manually annotated data is limited. An additional case study with mining-related articles reveals insights into the supply chain network, demonstrating the practical use of this method.

2 - Enhancing Supply Chain Resilience and Equitable Access: A Strategic Approach to the Infant Formula Shortage

Chun-Miin Chen, Bucknell University, Lewisburg, PA, United States

The 2022 infant formula shortage in the United States exposed significant vulnerabilities in the nation's supply chain, revealing the challenges of concentrated production among a limited number of manufacturers and a fragile distribution network. This crisis emphasized the pressing need for strategies to bolster supply chain resilience and ensure equitable access to essential products. This study leverages established approaches to address these challenges, focusing on distribution frameworks that tackle current weaknesses. The research explores various inventory management strategies and identifies key product availability factors, particularly in underserved regions. These strategies aim to enhance expected consumer surplus, improving overall consumer welfare. The findings highlight the importance of implementing targeted adjustments to distribution systems to strengthen supply chain resilience. This study contributes to the conversation on how best to redesign supply chains to enhance resilience and prioritize the needs of the most vulnerable populations.

3 - Resilience in the Semiconductor Supply Chain - A Systematic Literature Review

Zachary Ramirez, Purdue University, West Lafayette, IN, United States, Tho Le

Resilience has been a growing area of interest within supply chain research. This has been driven by the increasing frequency of disruptive events within the globalized supply chain such as pandemics, geopolitical conflicts, and extreme weather events. Furthermore, rising demand for advanced electronics places semiconductors in a critical position for ensuring global security and promoting innovation. This study seeks to assess current resilience levels in the semiconductor supply chain by analyzing historical disruptions and their impacts, identifying critical vulnerabilities and bottlenecks in the supply chain, and developing a comprehensive understanding of the factors contributing to supply chain disruptions through literature review of research on supply chain resilience in the semiconductor industry.

The review was conducted systematically following the PRISMA approach with papers related to semiconductor supply chain resilience and risk management sourced from Web of Science and Connectedpapers.com. The final selection included 45 papers spanning September 2001 to February 2024. The papers were then analyzed to determine sources of risk and mitigation strategies. This analysis concluded that the most discussed sources of risk in the semiconductor industry were hardware security and shortage effects. Following the analysis, a framework was proposed to model resilience in the semiconductor supply chain and assess the effectiveness of risk management options.

4 - The Interplay between Supplier’s Demand Forecast Sharing and Channel Encroachment

Wenxin Zhang, Syracuse University, Syracuse, NY, United States, Zhengping Wu

Motivated by the observation that many suppliers have their own private demand forecasts, we study the strategic interplay between encroachment and information sharing decisions using the signaling game framework, where the strategic interplay refers to different decision sequences: whether the supplier makes the encroachment decision *before or after* deciding whether to share his private information with the retailer. In addition, three information-sharing formats are examined, including no information sharing, voluntary information sharing, and mandatory information sharing. We characterize the equilibrium outcomes of the different scenarios, which yield the following managerial insights: First, when the supplier decides whether to encroach *before* sharing information, the results are consistent with the simultaneous decision-making situation. Specifically, whether the supplier chooses to encroach or not, information sharing always benefits both firms. Second, when the supplier decides whether to encroach *after* sharing information, the sharing formats will strategically impact the encroachment decision, which is imperfect information for the retailer. Specifically, in the voluntary-sharing format, when the supplier chooses not to share information and encroaches into retailing, the upward distortion of wholesale prices will intensify downstream competition and then lead to a win-lose outcome. However, if the supplier opts not to encroach, it will lead to a win-win outcome, and the supplier can earn the highest profit.

WA11

Summit - 331

Accelerated Computing for Mathematical Programming

Invited Session

OPT: Computational Optimization and Software

Chair: Sungho Shin, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: François Pacaud, Mines Paris - PSL, Paris, France

Co-Chair: Mihai Anitescu, Argonne National Laboratory, Lemont, IL, 60439, United States

1 - Nonlinear programming on GPU with MadNLP

François Pacaud, Mines Paris - PSL, Paris, France, Sungho Shin

In this talk, we explore an interior-point method for solving large-scale constrained nonlinear programs with the aid of graphics processing units (GPUs). The interior-point method involves resolving a sequence of symmetric indefinite linear systems, or Karush-Kuhn-Tucker (KKT) systems, which become increasingly ill-conditioned as we approach to the solution. Solving the KKT systems with traditional sparse factorization methods involve numerical pivoting that makes parallelization difficult. Instead, we reformulate the KKT systems into symmetric positive-definite systems using a condensed space approach. As a result, we can solve the KKT systems efficiently on the GPU using sparse Cholesky factorization. We implement the method on GPUs with MadNLP.jl, a solver interfaced both with the NVIDIA library cuDSS for solving sparse positive definite systems and ExaModels.jl for evaluating models on GPUs. We assess the performance on large-scale problems from the PGLIB and COPS benchmarks, revealing that GPUs can achieve up to a tenfold speed increase compared to CPUs.

2 - HiOp: An HPC Optimization Suite with Exa-Scale Computing Capability

Nai-Yuan Chiang, Lawrence Livermore National Laboratory, Livermore, CA, United States, Cosmin Petra, Jingyi Wang

HiOp is an optimization suite designed for tackling large-scale mathematical programming challenges efficiently. This lightweight High-Performance Computing (HPC) solver takes advantage of an application's existing data parallelism to parallelize optimization iterations using specialized parallel linear algebra kernels. Acceleration can be achieved through OpenMP, CUDA, or HIP, thanks to the RAJA portability abstraction and various execution policies. HiOp has demonstrated its capabilities in solving large-scale optimization problems on parallel machines equipped with AMD or NVIDIA accelerators. Notably, it has been successfully applied to solve Security-Constrained AC Optimal Power Flow problems for networks featuring 10,000 buses, 10 renewable energy forecast scenarios, and a staggering 100,000 contingencies on the Frontier platform, all achieved within an impressive 20-minute timeframe.

3 - Acceleration of (Nonlinear) Optimization for Robotics through Co-Design: from MCUs to GPUs

Brian Plancher, Barnard College, Columbia University, New York, NY, United States

Model predictive control (MPC) is a powerful tool for controlling highly dynamic robotic systems subject to complex constraints. However, MPC, and its underlying (nonlinear) optimization algorithms, are often too computationally demanding to meet real-time rates for robotic platforms, both large and small. By leveraging a combination of parallelism, approximation, and structure exploitation, we have co-designed accelerated (nonlinear) trajectory optimization solvers for real-time performance on non-standard computational hardware, ranging from microcontrollers (MCUs) to graphical processing units (GPUs). These results have enabled us to run MPC onboard an MCU powered 27g quadrotor, as well as to run whole-body nonlinear MPC at kHz rates for a GPU powered manipulator, demonstrating high-speed trajectory tracking and dynamic obstacle avoidance.

4 - Advances in direct linear solvers for enhancing interior methods on GPU

Alexis Montoison, GERAD, Montréal, QC, Canada, Iain Duff

The solution of sparse linear systems has been the major technical challenge for implementing continuous optimization solvers on GPUs. Due to this challenge, GPU implementations of optimization algorithms have been primarily based on algorithms that do not require the use of matrix factorization, such as first-order methods or iterative linear solvers.
However, recently there has been significant progress in utilizing sparse direct linear solvers within optimization solvers running on GPUs. This presentation discusses some of the recent progress in the direct sparse linear solver capabilities on GPUs.
First, we highlight some of the impressive capabilities of NVIDIA's cuDSS library when applied to optimization solvers.
In particular, the recent NVIDIA cuDSS toolbox offers LU, LDL', and LL' decompositions, significantly enhancing the computational efficiency of interior methods.
We introduce the Julia interface CUDSS.jl and demonstrate its integration with the nonlinear optimizer MadNLP.jl.
Second, we discuss developing a new GPU-accelerated linear solver, based on the mature implementation of MA57, to deliver scalable solutions for symmetric indefinite linear systems.
We will discuss the distinct challenges posed by the nature of GPUs and discuss the strategies to exploit the BLAS-like routines on GPUs to maximally utilize the parallel computing capabilities of GPUs.
We also discuss the efforts to ensure portability across different GPU architectures, including NVIDIA, AMD, and Intel.

5 - OSQP with GPUs & FPGAs: Accelerating quadratic programming on heterogeneous systems

Ian McInerney, Imperial College London, London, United Kingdom, Maolin Wang, Bartolomeo Stellato, Vineet Bansal, Amit Solomon

The rise of machine learning has led to increased availability of accelerators, including GPUs in workstations and cloud services like AWS offering GPUs and FPGAs. However, traditional optimization solvers have largely not leveraged this computational power. In this talk, we present our recent work in the OSQP (Operator Splitting Quadratic Programming) 1.0 release, which implements end-to-end solver acceleration on both GPU and FPGA platforms. At the core of this acceleration is a linear algebra abstraction layer that decouples the core optimization steps from numerical operations, enabling a unified OSQP solver API regardless of the computational backend. This layer allows for seamless backend selection at build-time, including CPU backends with sparse operations or Intel MKL, a GPU backend using cuSPARSE, and an FPGA backend using RSQP. OSQP libraries with these backends are packaged with high-level interfaces, such as the OSQP Julia and Python packages, enabling users to easily solve quadratic optimization problems on GPUs.

WA12

Summit - 332

Operations and Revenue Management for AI Development

Invited Session

Revenue Management and Pricing

Chair: Ningyuan Chen, Rotman School of Management, Toronto, ON, Canada

Co-Chair: Rafid Mahmood, University of Ottawa + NVIDIA, Ottawa, ON, Canada

1 - R&D Data Sharing in New Product Development

Jussi Keppo, National University of Singapore, Singapore, Singapore, Zhi Chen

Many innovations today are data-driven such as self-driving cars. To improve the algorithms of these products, firms make substantial investments in data collection. However, the data is limited for an individual firm. This paper studies whether data collected by individual firms should be shared in order to promote innovation.

2 - Operations for New ML Products by Optimally Stopping Data Collection

Rafid Mahmood, Telfer School of Management, Ottawa, ON, Canada, Yufeng Zheng, Ningyuan Chen

The proliferation of machine learning (ML) models as software products is limited by the escalating costs of model development. These costs lie primarily in the collection of training data, since more training data leads to better models and thus, better products. In this work, we explore the problem of optimizing the development and deployment of the ML software from a revenue management perspective. Here. a firm must forecast when the diminishing value of marginal improvements to their ML product with respect to user demand is outweighed by the costs of collecting the additional data needed to obtain these improvements. We estimate the neural data scaling laws associated with the model and apply a stopping policy that achieves bounded regret with respect to revenue obtained from user demand. Simultaneously, we explore the effect of this forecast on the prices that the firm can set for the ML product. Numerical evaluation on several ML tasks reveal that, rather than the current trends of perpetually chasing bigger datasets and better models, firms can significantly reduce their development costs by learning to stop early and deploying a “just good enough” ML model that meets user demand.

3 - When Emotion AI Meets Strategic Users

Yifan Yu, University of Texas at Austin, Austin, TX, United States, Wendao Xue, Lin Jia, Yong Tan

When organizations adopt artificial intelligence (AI) to recognize individuals' negative emotions and accordingly allocate limited resources, strategic users are incentivized to game the system by misrepresenting their emotions. The value of AI in automating such emotion-driven allocation may be undermined by gaming behavior, algorithmic noise in emotion detection, and the spillover effect of negative emotions. We develop a game-theoretical model to understand emotion AI adoption, particularly in customer care, and analyze the design of the associated allocation policies. We find that adopting emotion AI is valuable if the spillover effect of negative emotions is negligible compared to resource misallocation loss, regardless of algorithmic noise and gaming behavior. We also quantify the welfare impacts of emotion AI on the users, organization, and society. Notably, a stronger AI is not always socially desirable and regulation on emotion-driven allocation is needed. Finally, we characterize conditions under which leveraging the AI system is preferred to hiring human employees in emotion-driven allocation. We also explore the alternative application of using emotion AI to monitor strategic employees and compare it with hiring a human manager for monitoring. Intriguingly, algorithmic noise may increase the profit of AI monitoring. Our work provides implications for designing, adopting, and regulating emotion AI.

WA13

Summit - 333

Innovative Approaches to Inventory Management and Sustainable Sourcing

Contributed Session

Chair: Seongwon Park, Michigan State University, East Lansing, 48824

1 - Retailer Matching for the Environment in a Supply Chain

Ji Young You, National Taiwan University, Taipei, Taiwan, Chialin Chen

In recent years, there has been a significant reevaluation of the role of retailers in promoting environmental sustainability within supply chains. This shift is exemplified by initiatives from major players like Walmart and Alibaba, who actively collaborate with green manufacturers and consumers. Contrary to traditional views associating retailers with negative impacts on supply chain sustainability, contemporary evidence suggests otherwise. This study delves into how retailers can influence green supply chains, considering two scenarios: direct competition between traditional and green manufacturers, and a retailer-facilitated "matching" model.

In direct competition, eco-friendly options may be limited, but in the matching model, retailers connect consumers with the best-suited green products from various manufacturers, making green options readily available. This challenges the conventional perception of retailers hindering sustainability and highlights their potential to promote environmentally friendly products.

Our findings emphasize the pivotal role retailers can play in advancing environmental responsibility by enhancing the accessibility of green products through effective consumer-product matching. This underscores the necessity to reconsider the role of retailers in supply chain sustainability and underscores the potential benefits of collaborative efforts between retailers, manufacturers, and consumers in driving positive environmental outcomes.

Our research provides valuable insights into how retailers can leverage their influence to drive sustainability initiatives and contribute to a more environmentally conscious marketplace. By redefining their role and embracing collaboration, retailers can become key drivers of positive change in the pursuit of a greener future.

2 - The Role of Retail Competition in Store Brand Introduction Decisions

Alexander Maslov, University of Kentucky, Lexington, KY, United States, Yasin Alan, Mumin Kurtulus, Chunlin Wang

We study a supply chain with a single national brand manufacturer that sells to consumers through two competing retailers who have an option to introduce a store brand. We explore the role of competitive interactions between the two retailers and the national brand manufacturer on the store brand introduction decisions. We find that existing studies overestimate the marginal benefit of store brand introduction. In addition, we characterize conditions under which one retailer may benefit from the other retailer introducing a store brand. Finally, we show that competition may force retailers to introduce store brands even though it results in lower profits for both retailers.

3 - Dynamics of Environmental Management in Multi-tier Supply Chains: Focusing on Aggregate Effects

Seongwon Park, Michigan State University, East Lansing, MI, United States, Srinivas Talluri, Seetharama Chandrasekhar Manchiraju, Gyusuk Lee

To reduce adverse environmental impact in supply chains, multinational corporations not only manage their own operations internally but also collaborate externally with other members of their supply chains. Although firms are increasingly sharing their environmental performance with external stakeholders, empirical studies assessing the environmental impact of cross-tier interactions between member firms in multi-tier supply chains are scant. The primary goal of this study is to address the differential environmental impacts of downstream firms on the upstream partners in multi-tier supply chains. To achieve this, we draw upon supply chain relationships data provided by FactSet Revere and Trucost environmental performance data and create firm-level environmental efficiency scores, across the supply chains of 6 different manufacturing industries, using network data envelopment analysis. We use a 5-year panel data from 2014 to 2018 based on 259 open triadic relationships. Our empirical analysis demonstrates that a focal firm's environmental efficiency positively influences its first-tier supplier's environmental efficiency. We also find that this influence increases with the financial dependence of the first-tier supplier on the focal firm. Moreover, while the first-tier supplier's environmental efficiency positively influences the second-tier supplier's environmental efficiency, focal firms may not directly impact the environmental efficiency of second-tier suppliers.

4 - Location based omnichannel assortment planning

Qingwei Jin, Zhejiang University, Hangzhou, China, People's Republic of, Zhi Pei

We consider a location based omnichannel assortment planning problem. The retailer is running several offline and online stores in an area. Each online store order is fulfilled by nearby offline stores through delivery service. The assortment in each offline store is serving both offline and online stores and the service relation forms a network. The retailer needs to determine the optimal assortment in each offline store and each online store. We design a partition rule to divide the network and propose a 1/2-approximation scheme when the network is a long chain and 1/3-approximation scheme when the network is induced from a 2-dimensional plane.

5 - A framework for retailers and manufacturers to maximize their profits through the utilization of food shelf-life extension technology

Yue Yao, University of Tennessee, Knoxville, Knoxville, TN, United States, Mingzhou Jin

We are going to build a framework for retailers and manufacturers to maximize their profits through the utilization of food shelf-life extension technology. The retailer needs to decide inventory cycle length, selling price, and order quantity for given extended shelf-life and wholesale price to maximize the total profit after considering the loss caused by expiry. The manufacturer needs to decide the length of the extended shelf-life and wholesale price to maximize his profit.

WA14

Summit - 334

New Topics in Grocery Retailing

Contributed Session

Chair: Yasamin Salmani, Bryant University, Smithfield, RI, United States

1 - Strategic Prediction and Decision Model for Operational Plan in Multichannel Food and Grocery Retailing

Jingran Zhang, Marshall University, Huntington, WV, United States

The Online Food and Grocery retailing is an increasingly growing market of Grocery Delivery and Meal Delivery. Customers turned to e-commerce for all their daily needs during and post pandemic, including food and groceries. Retailers provide different ways to process and fulfill customer orders from both online and offline channels, but one of the biggest challenges is last mile delivery stage. As a result, they need to make strategic operational decisions on channel switch and maintain. In this study, an integrated strategic decision model is established with demand prediction and expense estimation of delivery, and quick response to consumer orders to optimize cost-cutting of retailers and order fulfillment performance.

2 - Learning Commissions and Subscription Fees under Uncertainty in Two-Sided Marketplaces

Christina Liepold, School of Management, Technical University of Munich, Munich, Germany, Maximilian Schiffer

Two-sided marketplaces are online platforms that facilitate exchanges between suppliers and buyers as they reduce search costs, for example, in the context of vacation rentals, freelancing, or salon bookings. These platforms charge marketplace fees, typically through commissions or subscription fees, which are crucial for revenue generation and long-term economic success of the marketplace.
Existing research approaches focus on deterministic settings, analytically evaluate optimal marketplace fees, and discuss how to optimally fix commissions and subscription fees a priori to maximize marketplace operator revenue or overall marketplace welfare. In practice, however, marketplace operators may face multiple uncertainties regarding the market setup that complicate the fee-setting procedure, such as suppliers' and buyers' stochastic arrivals, pricing expectations, and maximum waiting times. In this context, we study an adaptive design of marketplace fees based on deep reinforcement learning (DRL), enhancing the operator's ability to navigate inherent market uncertainties. Exploring the intersection of DRL and the platform economy, this approach allows for non-myopic fee-setting in marketplaces by considering the impact of these fees on the overall welfare of all marketplace participants and the operator's revenue.

3 - Batching of Grocery Orders with Attribute Selection in an Island Layout

Zijia Wang, New Jersey Institute of Technology, Newark, NJ, United States

The order picking literature focuses exclusively on specific or unique items. Grocery item order picking includes a class of items that are characterized by attributes associated with customer preferences. Fresh produce such as fruits and vegetables have natural variations, and each unit of the same item is different. Picking three apples from a pile of gala apples has multiple outcomes in terms of weight, color, and ripeness. Grocery stores want to provide BOPS customers with item-specific attributes that are then transmitted to the pickers. Attribute-based order picking can be challenging, and this is the first research to address the problem. Item pick times are no longer fixed but a function of the number and complexity of the attributes. We develop non-linear functional models to describe the picking times and develop correlations across multiple orders. These items are typically stocked in an island layout with perimeter access. We present both MIP and heuristic solutions to the order picklist problem. The objective is to create order batches that minimize the combined travel and picking times. Experimental results show that the heuristics provide high-quality solutions in negligible time.

4 - Analysis of Factors Influencing Waiting Time for Unloading Trucks at Fresh Produce Wholesale Markets

Hiroki Iizuka, Waseda University, Tokyo, Japan, Kotomichi Matsuno, Shunichi Ohmori, Shintaro Nishibe, Takahiro Ohno

Wholesale markets serve as crucial infrastructures in the supply of fresh food products. Specifically, in the distribution channels for fresh produce, there is a flow between individual farmers, agricultural cooperatives, wholesale markets (wholesalers/intermediaries), and retailers. Information dissemination is one of the key functions of wholesale markets, enabling smooth transactions through information sharing among these entities. However, the uncertainty of shipping information until the last moment makes it difficult to implement effective strategies through information sharing. As a result, there is a concentration of truck arrivals at wholesale markets during late-night hours, exceeding the market's capacity, which leads to long waiting times for market entry. This issue directly impacts the working conditions of drivers and market staff, highlighting the urgency for a solution.

Current efforts focus on developing cargo reservation systems and improving unloading procedures to reduce overall waiting time. However, the magnitude of the impact of these initiatives remains unclear and has not been connected to significant decision-making processes. Therefore, it is necessary to validate the waiting time reduction effects generated by each proposed improvement.

In this study, we aim to simulate the entire process from truck transportation to unloading at the wholesale market in order to quantitatively demonstrate the effects of each proposal on reducing waiting time. This will enable us to determine the optimal allocation of resources and encourage the implementation of improvement measures.

5 - Strategic Sales Channel Management in Times of Crisis

Yasamin Salmani, Bryant University, Smithfield, RI, United States, Yuyun Zhong, Yang Li

In light of pandemic-induced disruptions, this study investigates the impact of sales channel mixes on the revenue of 84 U.S. manufacturing companies. Analyzing data from the last quarters of 2019 and 2020 with association rule mining, the research highlights how different sales strategies and product types interrelate within these channels. The results demonstrate that third-party retailers contribute positively to revenue, whereas company-owned physical stores often result in losses. Online channels prove most effective when aligned with experiential products. The study discerns that while multiple channels may decrease revenue, focused channel approaches—single or dual—tend to increase it. These insights aid in refining sales channel management during crises and suggest areas for further research.

WA15

Summit - 335

Dynamic Models in Revenue Management

Invited Session

Revenue Management and Pricing

Chair: WEIYUAN LI, Cornell Tech, New York, NY, 10044, United States

1 - Prophet Inequalities with Cancellation Costs

Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States, Farbod Ekbatani, Pranav Nuti, Jan Vondrak

Most of the literature on online algorithms and sequential decision-making focuses on settings with "irrevocable decisions" where the algorithm's decision upon arrival of the new input is set in stone and can never change in the future. One canonical example is the classic prophet inequality problem, where realizations of a sequence of independent random variables X_1, X_2, ... with known distributions are drawn one by one and a decision maker decides when to stop and accept the arriving random variable, with the goal of maximizing the expected value of their pick. We consider "prophet inequalities with cancellation costs" in the linear cost setting, where after accepting a variable X_i, we can still discard X_i later and accept another variable X_j, at a cost of f times X_i. The goal is to maximize the expected net reward, which is the value of the final accepted variable minus the total buyback cost. Our main result is a full characterization of the optimal competitive ratio as a function of parameter f everywhere. In particular, when f>1, we show an optimal competitive ratio of (1+f)/(1+2f). Our upper-bound result is obtained by (i) a sequence of reductions to identify worst-case instances, (ii) using combinatorial optimization techniques involving LP duality, flows, and cuts, and (iii) finally reducing the characterization of the optimal competitive ratio to solving a particular differential equation and showing how to solve that differential equation. We complement this upper-bound by a matching lower-bound

2 - Approximation Schemes for Dynamic Pricing with Opaque Products

Jacob Feldman, Olin Business School, Clayton, MO, United States, Yukai Huang, Danny Segev, Levi DeValve

In this paper, we consider a multi-period, multi-product dynamic pricing problem in which each product is endowed with an exogenous starting inventory level, and there is the added complexity of an opaque selling option. That is, alongside traditional (transparent) products, the retailer or platform also has the option to create and price an opaque product, which corresponds to a dummy product comprised potentially of any subset of the displayed transparent products. In the event that a customer selects the opaque product, the platform has the freedom to choose any of the opaque product's constituents to satisfy this demand. All-in-all, we are left with a classical dynamic pricing problem with a twist, since the addition of the opaque selling option gives the platform an extra lever of flexibility to balance supply and demand. For this novel pricing setting, we develop a nuanced constant factor approximation scheme.

3 - Prophet Inequalities for Product Recommendation Under the Bounded Last-Choice Model

Pin Gao, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Guillermo Gallego, Shaoyu Wang

This paper introduces a novel regular choice model that merely imposes an upper bound on each product's last-choice probability (i.e., the purchase probability of being exclusively recommended). Notable instances of this model include the Multinomial Logit (MNL), click-based MNL, sequential click-based MNL, and the general attraction model. For the static cardinality-constrained revenue-maximizing recommendation problem, despite its NP-hardness even under specific model instances, we propose a polynomial-time solvable constant-factor approximation heuristic for the general case, leveraging solely the last-choice probabilities. Additionally, we investigate an online recommendation scenario inspired by the burgeoning practice of live-streaming e-commerce. In this context, product-specific parameters are sequentially revealed as random and independent variables, compelling decision-makers to make immediate and irreversible decisions regarding whether to recommend a product upon receiving its information. We offer constant-factor approximations for cases where the information arrival order is adversarial or adheres to specific distributional patterns. Our comparative performance analysis underscores the importance of information in the online setting and establishes a novel connection between the online recommendation problem and the classical multiple-unit prophet inequality.

4 - Revenue Management with Calendar-Aware and Dependent Demands: Asymptotically Tight Fluid Approximations

Weiyuan Li, Cornell University, New York, NY, United States, Paat Rusmevichientong, Huseyin Topaloglu

When modeling the demand in revenue management systems, a natural approach is to focus on a canonical interval of time, such as a week, so that we forecast the demand over each week in the selling horizon. Ideally, we would like to use random variables with general distributions to model the demand over each week. The current demand can give a signal for the future demand, so we also would like to capture the dependence between the demands over different weeks. Prevalent demand models in the literature, which are based on a discrete-time approximation to a Poisson process, are not compatible with these needs. In this talk, we focus on revenue management models that are compatible with a natural approach for forecasting the demand. Building such models through dynamic programming is not difficult. We divide the selling horizon into multiple stages, each stage being a canonical interval of time on the calendar. We have random number of customer arrivals in each stage, whose distribution is arbitrary and depends on the number of arrivals in the previous stage. The question we seek to answer is the form of the corresponding fluid approximation. We give the correct fluid approximation in the sense that it yields asymptotically optimal policies. The form of our fluid approximation is surprising as its constraints use conditional expected capacity consumption. As the resource capacities and number of stages increase with the same rate, our performance guarantee converges to one.

WA16

Summit - 336

Network Design/Revenue Management and Platform Operations

Invited Session

Revenue Management and Pricing

Chair: Victor Araman, American University of Beirut, Beirut, Lebanon

Co-Chair: Heng Zhang, Arizona State University, Tempe, AZ, United States

Co-Chair: Yunduan Lin, The Chinese University of Hong Kong, Santa Clara

1 - Expectation vs. Certainty: The Influence of Discount Announcements on Purchasing Behavior

Frederico Zimmerman, Harvard Business School, Cambridge, MA, United States, Gustavo Vulcano, Joaquín Navajas

We conducted a study to examine whether pre-announcing discounts affects consumers' decisions to wait or buy in a simulated retail scenario. Contrary to rational models, we hypothesized that preannounced discounts would lead more individuals to wait for markdowns. Participants were placed in conditions with varying stock quantities and discount information, including scenarios with no discount preannouncements and specific preannounced discounts of 20%, 50%, and 80%. Our results consistently showed that preannounced discounts led to more participants waiting for markdowns. We also observed that participants in the non pre-announced condition who imagined larger discounts were also more likely to wait. Remarkably, even participants in the non-preannounced condition who expected a 20% discount were significantly more likely to wait for the markdown than those who were certain about a 20% discount in the preannounced condition. Overall, these findings suggest that not preannouncing discounts can increase consumer uncertainty and elicit more impulsive purchasing, which may be more beneficial for retailers.

2 - Analysis of a Competitive Distributed Supply Chain

Victor Araman, American University of Beirut, Beirut, Lebanon

We consider a two-tier supply chain model constituted of a retailer facing end-consumers' demand and supplied by a large number of small suppliers. The retailer selects an adequate base-stock policy while suppliers decide, each independently, on their capacity levels. We adopt a queueing theoretic-approach for the supply function and reduce the dynamics complexity through the use of mean field equilibrium. We show the existence of such equilibrium and obtain a full characterization of the solution. We suggest a mechanism design that guarantees full coordination of the supply chain.

3 - Diversified Learning: Bayesian Control with Multiple Biased Information Sources

Xinyuan Zhang, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada, Jussi Keppo, Michael Jong Kim

We consider a decision-maker (DM) who can acquire signals from multiple biased information sources to learn about a hidden state prior to making an earning decision. Unbiased signals are also available, but the acquisition cost is high. The DM jointly optimizes his learning (acquisition) and earning decisions to maximize the expected payoffs. This is particularly relevant in healthcare and revenue management, where decisions rely on data from different sources (e.g., combining diagnosis results, aggregating demand forecasts). We develop a Bayesian control framework for such problems, where multi-source learning is modelled by a hierarchical Bayesian network. Focusing on the budget-constrained scenario, we identify an optimal acquisition policy that promotes diversification across biased sources to mitigate misinformation risks. To support our theoretical findings, we also perform empirical and simulation studies on disease prevalence and consumer demand datasets. These studies underscore the vital importance of a diversified information acquisition strategy in complex decision-making scenarios.

4 - Information Disclosure and Consumer Search on the Online Experience Goods Platform: Theory and Empirical Evidence

Yushu Zeng, Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Ying-Ju Chen, Xin Wang, Hu Huang

This paper explores the vertical and horizontal quality information disclosure strategies in a scenario where firms sell experience goods to consumers through an online platform. In the theoretical part, We develop a stylized game-theoretical model to capture the equilibrium patterns. We then test our predictions using data from a web novel platform and find supporting real-life evidence in the empirical part. Our findings indicate that when the search cost is negligible and vertical quality information is private, firms are less inclined to disclose horizontal attribute information. Additionally, we observe that firms with lower vertical quality are more likely to disclose fit information compared to high-type firms. Furthermore, platforms exhibit a lack of incentive to provide consumers with vertical quality information when the search cost is low.

5 - E-Commerce Subset Selection Problems and Approximate Submodularity

Levi DeValve, University of Chicago, Chicago, IL, United States, Duygu Soylemez

Many challenging problems in managing e-commerce fulfillment networks can be posed as subset selection problems: which fulfillment arcs to select during the network design phase, where to place each SKU in the network during the inventory planning phase, and where to fulfill orders from during the execution phase. These impactful problems are computationally challenging due to their combinatorial structure, and both the academic and practitioner communities have recognized the need to develop effective heuristics. The academic literature has long understood submodularity to be an invaluable structural property for analyzing subset selection problems, for the intuitive reason that local changes in the objective can be used to bound global changes. Unfortunately, most subset selection problems arising in the e-commerce setting lack submodularity, due to the inherent “two-sided” nature of the supply and demand networks under consideration. We overcome this technical challenge using a novel form of approximate submodularity to analyze local search heuristics for these problems, proposing a general framework that provides new constant factor approximation guarantees.

WA17

Summit - 337

New Applications and Methodology in Pricing and Inventory Management

Invited Session

Revenue Management and Pricing

Chair: Meichun Lin, Singapore Management University, Singapore, Singapore

1 - Combatting Food Waste via Joint Pricing and Perishable Inventory Optimization

ZICHUN LIU, McGill University, Montreal, QC, Canada, Sentao Miao, Wei Qi

We address the simultaneous determination of pricing and inventory control for perishable food. The optimal policy is computationally intractable due to the curse of dimensionality. We then develop heuristic policy and prove asymptotic optimality under several parameter regimes, and show theoretical and numerical performance of the policy.

2 - Mostly Beneficial Clustering: Aggregating Data for Operational Decision Making

Chengzhang Li, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhenkang Peng, Ying Rong

With increasingly volatile market conditions and rapid product innovations, operational decision-making for large-scale systems entails solving thousands of problems with limited data. Data aggregation is proposed to combine the data across problems to improve the decisions obtained by solving those problems individually. We propose a novel cluster-based Shrunken-SAA approach that can exploit the cluster structure among problems when implementing the data aggregation approaches. We prove that, as the number of problems grows, leveraging the given cluster structure among problems yields additional benefits over the data aggregation approaches that neglect such structure. When the cluster structure is unknown, we show that unveiling the cluster structure, even at the cost of a few data points, can be beneficial, especially when the distance between clusters of problems is substantial. Our proposed approach can be extended to general cost functions under mild conditions. When the number of problems gets large, the optimality gap of our proposed approach decreases exponentially in the distance between the clusters. We explore the performance of the proposed approach through the application of managing newsvendor systems via numerical experiments. We investigate the impacts of distance metrics between problem instances on the performance of the cluster-based Shrunken-SAA approach with synthetic data. We further validate our proposed approach with real data and highlight the advantages of cluster-based data aggregation, especially in the small-data large-scale regime, compared to the existing approaches.

3 - Online Learning for Dynamic Pricing in Consumer Electronics Trade-in Program

Sean Zhou, Chinese University of Hong Kong, New Territories, Hong Kong, Zhuoluo Zhang, Wenhao Li

We consider joint learning and pricing for an electronics trade-in platform which buys and sells multiple used electronics. The platform offers both trade-in-for-cash and trade-in-for-upgrade options and sets the corresponding prices to acquire the used products and their reselling prices after simple cleaning and refurbishing. Both supply and demand models are parametric. We propose two heuristics and analyze their regret bounds.

4 - Balancing Between Sustainability and Profitability on Online Food Platforms

Yeqing Zhou, Eindhoven University of Technology, Eindhoven, Netherlands, Salma Abououkal, Nourhan Sakr

In an era where food waste is a growing concern, we model an online platform that matches unsold surplus food from various outlets with customers willing to order 'surprise boxes.' The inherent unpredictability of leftover inventory and customer choices often results in either excess waste or unmet demand, impacting both sustainability and profitability. To address this challenge, we propose dynamic algorithms that recommend a subset of stores to customers based on sales data. This targeted approach not only promotes a more equitable distribution of surplus food but also enhances the platform’s operational efficiency. Preliminary results demonstrate a significant reduction in food waste along with relative revenue stability. Our paper details the proposed algorithmic solutions, evaluates their effectiveness through simulation, and discusses the implications for both food sustainability and business performance.

WA18

Summit - 338

Empirical and Behavioral Insights in Revenue Management and Pricing

Invited Session

Revenue Management and Pricing

Chair: Freddy Lim, National University of Singapore, Singapore, Singapore

Co-Chair: So Yeon CHUN, INSEAD, France, Singapore

1 - Coins, Cards, OR Apps: Impact of Payment Methods on Street Parking Occupancy and Search TIMES

Sena Onen Oz, McGill University, Montreal, QC, Canada, Mehmet Gumus, Wei Qi

City dwellers often struggle with on-street parking. In most cities, once finally finding a curbside spot, drivers must pay in advance, but estimating the exact parking duration needed can be challenging. However, mobile payment applications enable drivers to adjust their parking duration remotely, if necessary, unlike traditional payment methods such as cash and credit cards. This paper examines how different payment methods and hourly parking prices affect drivers’ parking behavior, street parking occupancy, and search time. We investigate the influence of payment methods by conducting an online survey wherein participants make payment decisions under various conditions. Additionally, in collaboration with a municipal agency from a densely populated North American city, we examine high-resolution, multi-year records of street parking payments to evaluate the effect of pricing on payments. Our findings indicate that drivers' payment amounts vary significantly depending on their preferred payment methods and prices. Specifically, drivers tend to pay less using mobile applications than credit cards and cash, and decreasing hourly prices increases street parking payments. To provide further guidance to municipalities, we simulate parking scenarios to examine the effects of these factors on occupancy and search times. Our findings revealed that decreasing prices significantly increases the search time and street occupancy. We also found that the progressively increasing pricing policy significantly impacts both operational metrics compared to constant pricing. Lastly, we showed that mobile payment adoption could reduce search times and occupancy thanks to the reduced overpayment of mobile payers.

2 - (Machine) Learning Consumption Preferences in Service Networks

Kenneth Moon, University of Pennsylvania, The Wharton School, Philadelphia, PA, United States

The shopping paths that time-conscious customers take through service networks (e.g., supermarkets and shopping centers) reveal their choices over multiple stations with different offerings (e.g., a supermarket's sections for distinct categories), yet data on such paths present unique challenges for empiricists. Customers often realize value by visiting specific groups of complementary stations, and they navigate between the combinatorially many possible network paths to reach their destinations most conveniently and quickly. When using such data to recover customers' underlying consumption preferences and patience for time spent shopping, existing empirical methods exhibit surprising limitations, including intractability or inconsistency. We develop new estimators to address several major issues and apply them to an empirical case study utilizing customer mobile location data from a hypermarket serving over 1M shoppers annually. Despite the combinatorial complexity of network choice sets, we show that making a small set of utility comparisons against ``local'' alternatives suffices to validate that a customer is visiting her optimal set of stations and, for empirical researchers, to maximally learn customers' underlying preferences from observing their chosen paths in data. Neural networks capably exploit such hidden low-dimensional structure, which motivates us to develop estimators that use them to discriminate between and thus identify different customer types from shopping paths data.Our empirical methods enable analyses of service network management and customer targeting that account for station complementarities and customer patience. For a simple network staffing problem at the hypermarket, we find

3 - Points Plus Money: Uncovering the Impact of Mixed Currency Payments

Freddy Lim, INSEAD, -, Singapore, So Yeon CHUN

We explore the impact of implementing a mixed currency payment method, which allows consumers to use a combination of points and money for payments, on consumer behaviors and firm revenues. We develop a two stage self-selection model with instruments to empirically investigate the adoption of this payment method and its post-adoption outcomes within a large scale coalition loyalty program. After controlling for self-selection, the findings reveal that mixed currency payments increase consumers' point earning, point spending, and monetary spending. This change in consumer behavior boosts firm revenue, and we identify that a consumer segment often overlooked in marketing efforts contributed significantly to this revenue increase. Finally, we discuss the potential mechanisms behind this effect.

4 - Repair OR Replace: An Empirical Analysis of the Impact of Diagnostic Decisions on Product Returns in a Bike-Sharing Platform

Jingxuan Geng, Temple University, Philadelphia, PA, United States, Hailong Cui, Guangwen Kong, Guo Li

Bike-sharing has emerged as an effective solution for consumers in recent years. However, the growing consumer adoption of bike-sharing has led to an increase in the susceptibility of shared bikes to frequent damages, resulting in escalated operational costs. Our study focuses on the maintenance practices employed by bike-sharing platforms to address this challenge. We have gathered a comprehensive operational dataset from a prominent bike-sharing platform in Asia. In our reduced-form analysis, we first establish that decisions regarding the repair or replacement of defective bike components significantly influence the likelihood of the bike being returned for maintenance in a short timeframe. Subsequently, we delve into the development of a structural estimation model to examine how repair workers make diagnostic decisions and how these decisions impact the overall maintenance expenses for the platform.

WA19

Summit - 339

Pricing and Market effects

Contributed Session

Chair: Vicente Ahumada, Universidad Adolfo Ibáñez, Viña del Mar, 2540050, Chile

1 - Fairness-aware Contextual Dynamic Pricing with Strategic Buyers

Pangpang Liu, Purdue University, West Lafayette, IN, United States, Wei Sun

Contextual pricing strategies are prevalent in online retailing, where the seller adjusts prices based on products' attributes and buyers' characteristics. While such strategies can enhance seller's profits, they raise concerns about fairness when significant price disparities emerge among specific groups, such as gender or race. These disparities can lead to adverse perceptions of fairness among buyers and may even violate the regulation and law. Conversely, price differentials can incentivize disadvantaged buyers to strategically manipulate their group identity to pursue a lower price. In this paper, we investigate the contextual dynamic pricing with fairness constraints, taking into account buyers' strategic behaviors when their group status is private and unobservable from the seller. We propose a dynamic pricing policy that simultaneously achieves price fairness and discourages strategic behaviors. Our policy achieves an upper bound of $O(\sqrt{T})$ regret over T time horizons. We also prove an $\Omega(\sqrt{T})$ regret lower bound of any pricing policy under our problem setting. We support our findings with extensive experimental evidence, showcasing our policy's effectiveness. In our real data analysis, we observe the existence of price discrimination against race in the loan application even after accounting for other contextual information. Our proposed pricing policy demonstrates a significant improvement, achieving 34.75% reduction in regret compared to the benchmark policy.

2 - Bundling vs Two-part Pricing under Heterogeneity in Demand Uncertainty

Sreya Kolay, University at Albany, SUNY, Albany, NY, United States, Rajeev Tyagi

Two-part pricing and bundling are two commonly observed price discrimination strategies used by firms. For example, sellers such as Costco use two-part pricing through annual club membership fees in addition to per-unit prices. On the other hand, sellers such as Sling TV offer programing through bundles of different sizes. Prior academic literature has shown that bundling yields a higher profit compared to two-part pricing for a monopolist firm selling to consumers with heterogeneous valuations. This paper extends that research by introducing consumer uncertainty about their future demand needs. We show that under certain conditions on degree of consumer uncertainty and degree of consumer heterogeneity, two-part pricing can yield the seller a higher profit than under bundling. We also examine the implications of these pricing schemes on consumer welfare under consumer demand uncertainty.

3 - Data-driven advertising and pricing policy with budget constraint on e-commerce platform

Yunzhi Cao, City University of Hong Kong, Hong Kong, Hong Kong, Houmin Yan

Advertising and pricing are crucial tools for retailers on the e-commerce platform for promotion, such as Amazon. This study focuses on the joint bidding and pricing policy for retailers subject to budget limitation under multiple periods. The bidding is for the impression on the sponsored advertising list. The click-through rate and conversion rate functions with respect to bid and price are learned and estimated based on historical data. We formulate a stochastic model and construct a dynamic programming formulation. We prove the existence of the unique optimal solution and characterize the structure of an optimized bidding and pricing policy. Through theoretical analysis and a case study, we investigate the relationships among the optimal policy, value function, and key factors, and demonstrate the effectiveness of the optimal policy and the proposed dynamic model.

4 - Constraint learning: design and comparison of methods for the construction of constraints in multiproduct pricing

Vicente Ahumada, Universidad Adolfo Ibáñez, Viña del Mar, Chile, Luis Aburto

Optimizing multiproduct pricing is a challenging process. Demand models calibrated with transactional data tend to exhibit bias in cross elasticity effects, mainly due to endogeneity issues, leading to nonsensical or extreme outcomes during optimization. We examine different supervised and unsupervised methods to extract constraints from the data to formulate a robust feasible space for the pricing optimization process. These constraints, or pricing rules, reveal latent business rules that show the relationships between product prices. Moreover, they help to identify price solutions correlated with strong product category performance. We apply our methodology to an orange juice dataset from different supermarket stores, achieving significant and robust improvements in expected category profit, ranging from a 15% to 40% gain in profits. The methods used to create constraints for an improved optimization included an exhaustive parameter iteration, skope rules classifiers, robust optimization, definition of ellipsoidal feasible regions, support vector machines and tree classifiers.

5 - An Interpretable Auto-ML Pipeline for B2B Pricing Decisions

Zijian Zeng, PROS, Houston, TX, United States, Jonas Rauch, Manu Chaudhary, Mohit Mahajan, Chih-Hsien Wu, Royce Kallesen

In business-to-business pricing, organizations aim to maximize profits over time by developing optimized pricing strategies based on historical transactions involving a wide variety of customers, products, and other characteristics. Each customer receives a unique price for the product determined through negotiation, which makes it essential to consider both customer-specific prices (CSP) and market prices (MP) in pricing decisions. This task has become increasingly challenging in the era of big data due to the presence of millions of transaction records and countless possible combinations of transaction features, making it difficult for humans to analyze and optimize pricing effectively.

Current pricing solutions often require segmenting transactions to reduce dimensionality and define market boundaries. However, this approach limits cross-segment information sharing. To leverage all available data, we propose an automated machine learning (Auto-ML) pipeline for making smart pricing decisions. Our pipeline infers CSP and MP directly from transactions data without requiring segmentation, sacrificing interpretability, or requiring extensive manual tuning. It employs a deep-and-wide neural network to comprehensively explores feature information and cross-effects. Time-dependent bases can be added to identify linear and seasonal patterns at both market and transaction levels, if present. The pipeline uses a specialized hyperparameter search space to narrow down tunable parameters, enabling efficient Bayesian optimization for automatic tuning. At the end of training, Shapley values explain feature importance and identify the price changes attributable to each feature.

This Auto-ML pipeline has been used for daily pricing decisions in practice, delivering high credibility and strong results.

WA20

Summit - 340

Emerging Topics in Decision Analysis

Invited Session

Decision Analysis Society

Chair: Xiao Xiao, California State University, Stanislaus, Turlock, CA, United States

1 - Loyalty programs: when are they profitable for franchisees in hotel industry

Yunke Li, Tongji University, Shanghai, China, People's Republic of, Harihara Natarajan, Nan Yang

To scrutinize the impacts of loyalty programs, in this study, we develop a stylized game model featuring two firms with vertical product differentiation and a limited supply of products. Considering the fact that capacity commitment can soften price competition, and given the fixed number of rooms in hotels, our model's unique feature is that while it assumes that hotels cannot directly alter capacity, it does permit hotels to commit to a new room capacity by allocating a portion of rooms for reward redemption. Implementing reward redemption through loyalty program channel necessitates specific pricing and capacity constraints. We find that the hotel can enhance its profitability under specific conditions related to the proportion of loyal customers and redemption point requirements.

2 - Future of Healthcare: A New Perspective on Emerging Technology Adoption and Organizational Performance

Heng (John) Xie, California State University, Sacramento, Sacramento, CA, United States, Xinyu Wei, Richard Peng, Victor Prybutok

Rapid technological advancements provide healthcare organizations with increasing opportunities to enhance organizational performance. However, healthcare organizations face challenges in effectively understanding and leveraging these technologies. Previous studies have outlined the process through which users understand, accept, and utilize emerging technology, yet few have explored the combined effects of emerging technology adoption on hospital organizational performance. This research proposes a comprehensive model to study the impact of emerging technology adoption and integration on organizational performance in the healthcare industry. It specifically aims to integrate both upstream factors influencing technology adoption and downstream factors involved in measuring performance resulting from the utilization of these technologies. The results illustrate how emerging technology adoption can be effectively connected to hospital organizational performance. Additionally, this study discusses various factors influencing the adoption of emerging technologies by healthcare organizations and offers robust empirical evidence that can guide future research and inform decision-making processes within healthcare management.

3 - Managing Long-Term Supply Risks: Incentives for Smooth Production Flow

Xiao Xiao, California State University, Stanislaus, Turlock, CA, United States, Dawei Jian, Mohsen Elhafsi

Smooth production flow, a key aspect of lean manufacturing, is essential for firms to ensure continuous supply, stabilize prices, and maintain customer satisfaction. However, it is vulnerable due to dynamic supply risks. How should the firm incentivize a smooth production flow through procurement? We study a new class of long-term procurement contracting problem within a two-echelon supply chain, where a risk-averse buyer with market pricing power replenishes from a supplier whose private state of production is subject to random shocks and evolves dynamically over time. We derive the optimal procurement contract; it distorts order quantity in the initial period for screening purpose and subsequently intensifies quantities to align with production needs. We demonstrate that smooth production flow concern can reduce quantity distortion, addressing the agency problem and promoting a more stable supply chain. This work advances our understanding of addressing dynamic supply risks to foster smooth production flow.

4 - Invalid Extra Online Shopping for Free Shipping Policy

Si Liu, Kent State University, Kent, OH, United States

Delivery activity costs have become a significant concern for e-commerce retailers (ECRs). To strategically mitigate these costs, many ECRs impose criteria such as minimum purchase values or a specific number of items for customers to qualify for free forward shipping. However, these criteria are often absent for free returns (backward shipping), provided returns occur within a designated timeframe. This discrepancy in shipping policies can give rise to a phenomenon we term 'Invalid Extra Shopping' (IES), where customers purchase additional items with the intent to return them, solely to benefit from the free shipping policy that they would not otherwise qualify for. IES leads to revenue losses for ECRs and exacerbates delivery-related impacts such as congestion and pollution.

This paper addresses the IES phenomenon through a threefold approach. First, we develop a model to capture customer behavior associated with IES. Second, we analyze the influence of customer characteristics, purchase item classification, and ECRs' return windows on their revenues, identifying which ECRs are most susceptible to IES. Third, we evaluate the effects of various shipping and return policies on IES and propose optimal policies that benefit all stakeholders. Our findings offer critical insights for ECRs aiming to balance cost-effective operations with customer satisfaction.

WA21

Summit - 341

Informing AI Ethics

Panel Session

Decision Analysis Society

Co-Chair: Jacqueline Csonka-Peeren, University of Waterloo (alumni), Toronto, ON, Canada

1 - Panelist

Jacqueline Csonka-Peeren, University of Waterloo (alumni), Toronto, ON, Canada

2 - Panelist

Anajana Susarla, Michigan State University Broad College of Business, East Lansing, MI, United States

3 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

4 - Panelist

Amy Cohn, University of Michigan, Ann Arbor, MI, United States

5 - Panelist

Julie Ivy, University of Michigan, Ann Arbor, MI, United States

WA22

Summit - 342

Stochastic Analysis of Energy Power Systems

Contributed Session

Chair: Joonho Bae, University of Michigan Ross School of Business, 2301 Stone Rd, Ann Arbor, MI, 48105, United States

1 - Assessing the Resilience of Integrated Energy Systems under Capacity Expansion Risks

Weijie Pan, The George Washington University, Washington, DC, United States, Ekundayo Shittu

Within the framework of sustainable development, the decarbonization of electricity generating systems plays a pivotal role in climate change mitigation strategies. Energy policies have led to increases in the integration of renewable energy (REN) especially in capacity expansions. However, the enactment of these policies involves multiple considerations such as technical performance, economic effectiveness, and resource accessibility, all accompanied by significant uncertainties. While past literature extensively explores various expansion portfolios of REN technologies through qualitative discussions and quantitative modeling, few studies verify the feasibility of these policies in terms of physical system operations. Moreover, contemporary energy systems face high-impact, low-probability (HILP) threats from extreme environmental incidents and malicious human activities, necessitating enhancements in resilience. Consequently, the extent to which increased REN technology penetration affects system operations and resilience remains unclear. This study uses Modelica modeling language to capture the physics of energy systems in the presence of both HILP events and heightened REN generation technology penetration. Preliminary results illustrate the effectiveness of expanded REN capacities in recovering post-disaster system performance, contingent upon an assessment of the generation mix in the presence of the intermittency introduced by REN resources. By assessing the impact of energy policies on capacity expansion, this study contributes to informing how decision-making for system resilience enhancement is influenced by the physics of technologies.

2 - The influence of additionality and time-matching requirements on the emissions from grid-connected hydrogen production

Dharik Mallapragada, Chemical and Biomolecular Engineering, Tandon School of Engineering, New York University, New York, NY, United States, Michael Giovanniello

Decarbonization policies emphasize electrifying end-uses while growing low-carbon electric power supply. This approach raises the question of the induced emissions attributable to a specific load connected to the grid that also contracts variable renewable energy (VRE) power supply. The answer is not obvious because instantaneous power flows from a particular producer cannot be directly associated with a particular user. Nonetheless, such assessments are necessary and in the case of electrolytic hydrogen (H_2) production, have billion-dollar ramifications due to the emission-indexed H_2 production tax credit (PTC) in the U.S.. There has been a vigorous debate on the qualifying requirements for grid-connected H_2 production to receive the PTC, with the focus on: a) requirements on “time-matching”, i.e. the period over which electricity use must match production from contracted generators, and b) “additionality”, which require contracting newly built VRE plants. Here, we use an open-source macro-energy system model, DOLPHYN, to show that the emissions and cost impact of alternative time-matching requirements are conditional upon the way additionality is modeled. Significantly lower emissions are achievable under annual time-matching in the additionality framework that presumes VRE for non-H_2 electricity demand does not compete with VRE for H_2 (“non-compete” framework), as opposed to the framework where all VRE resources are in direct competition (“compete”). In practice, the “non-compete” framework can be realized with binding and stringent grid decarbonization targets. This implies that for regions with such policies, less stringent (and therefore less costly) time-matching requirements are sufficient to avoid system-level emissions increases due electrolytic hydrogen deployment.

3 - Assessing the Vulnerability of Power Systems in Decarbonized Power Grids: A Network Interdiction Study

Veronica Veronica Diaz, North Carolina State University, Raleigh, NC, United States, Jordan Kern

Decarbonizing electric power systems will likely involve greater reliance on variable renewable energy, especially wind, and solar, which provide lower electricity output per land area than fossil fuel and nuclear-based power. As a result, a distinct feature of future bulk power systems may be more numerous, lower-capacity generators spread more evenly throughout the network. The impacts of this new configuration on the vulnerability of system operations are still unclear. Motivated by recent physical attacks on grid infrastructure, this study examines how decarbonization through variable renewables could alter the vulnerability of the power system to intentional attacks. We explore this question as a network interdiction problem, a two-player sequential game model where an adversary aims to maximally damage a cost-minimizing system operator by destroying grid components (generators, transmission lines, substations, etc.). While network interdiction has previously been used to assess the vulnerability of power systems, its application in the context of system-wide decarbonization is novel. We use a clustering procedure on annual load and renewable generation data from a real-world 662-bus regional transmission system to generate representative operating scenarios. We then evaluate how an attacker's optimal strategies change throughout a typical operating year, contrasting the vulnerabilities of the current and future (decarbonized) grid configuration, including the capacity mix and its distribution throughout the regional transmission network.

4 - Navigating Challenges arising from Large-scale Renewable Energy Integration and Solutions to Address Them

Elnaz Kabir, Texas A&M University, College Station, TX, United States

The global priority to combat climate change has led to a focus on decarbonizing power systems. However, the variability and intermittency of renewable energy sources pose several challenges for power systems operations, including energy curtailment, load shedding, and electricity price volatility. In this presentation, we analyze these challenges and propose solutions to address them using a real world case study.

5 - Cost-Saving Synergy: Energy Stacking in Battery Energy Storage Systems

Joonho Bae, University of Michigan Ross School of Business, Ann Arbor, MI, United States, Roman Kapuscinski, John Silberholz

Despite the great potential benefits of battery energy storage systems (BESS) to electrical grids, most standalone uses of BESS are not economical due to batteries' high upfront costs and limited lifespans. Energy stacking, a strategy of providing two or more services with a single BESS, has been of great interest to improve profitability. However, some key questions, e.g., the underlying mechanism by which stacking works, or why and how much it may improve profitability, remain unanswered in the literature. Using two popular battery services, we analytically show that there often exists cost-saving synergy -- the cost of performing both services at the same time (simultaneous stacking) is smaller than the sum of individual costs if we had performed each service alone -- which allows for bigger profits. Furthermore, we perform comparative statics on the optimal mix of the services to systemically characterize grid/market conditions that maximize/minimize this synergy. We also derive a theoretical upper bound on simultaneous stacking's benefits, showing that it can approximately double the profit of the best standalone service. Several generalizations of the base model not only show that the main lessons continue to hold but also that stacking's benefits may become even stronger.

6 - Optimizing Energy Consumption in AI Training: A Reinforcement Learning Approach

Viveka Saraiya, Independent Researcher, Seattle, WA, United States, Mayuree Binjolkar

With artificial intelligence (AI) developing at a rapid pace, and with its growth potential increasing by the day, the question arises of what amount of energy is required to support this growth. The energy efficiency of an AI model is an emerging research area that needs to be explored further in order to understand any potential implications of this growth. Based on this motivation, the purpose of the study is to examine opportunities to more efficiently utilize the energy required to train an AI model. We develop a dynamic predictive model that uses reinforcement learning to continuously adjust optimization strategies to reduce its energy consumption. The primary constraints are to minimize gains in training time and to maintain the original range of model accuracy.The model evaluates making changes to hardware setups, software settings, and training methods based on continuous model performance feedback. The results show that with the optimized combinations of rewards and robust model formulation, energy efficiency of AI can be maintained while making incremental changes to the different setups.

WA23

Summit - 343

AI and Technology for Humanitarian Logistics and Supply Chain Resilience

Contributed Session

Chair: Ahana Malhotra, McMaster University, Hamilton, ON, L8P1C3, Canada

1 - Using Machine Learning to Predict Late Deliveries in International Healthcare Logistics

Nima Molavi, California State University, San Bernardino, San Bernardino, CA, United States, Jeevan Sai Gali

The global healthcare supply chain encounters numerous challenges, ranging from natural disasters to logistical inefficiencies, which can lead to disruptions in the delivery of essential medical commodities. Machine learning models can be applied to the healthcare big data to predict and mitigate late deliveries, and consequently ensure timely access to healthcare resources. This research aims to contribute to the body of knowledge by (1) Identifying influential factors in late deliveries in international healthcare logistics, and (2) Developing Machine Learning models based on the identified factors to predict late deliveries in international healthcare logistics.

The identified factors in this paper can be listed under three major categories: Transportation Characteristics, Shipment Characteristics, and Country Capabilities. To study the impact of these factors on late deliveries, the USAID Health Commodity Delivery Dataset is modified and complemented by the Logistics Performance Index (LPI) to generate a more holistic view of the problem. Next, to predict the delay in shipments, several Machine Learning models such as Neural Networks, Random Forests, and Support Vector Machines, are developed based on the identified influential factors. The performance of these models have been assessed and compared to provide additional insights. To sum it up, by understanding the factors contributing to disruptions, exploiting advanced analytics to predict late deliveries, and developing proactive strategies for the industry, this research strives to improve the efficiency and reliability of healthcare commodity logistics worldwide.

2 - Technology Trends in Human Trafficking

Wojtek Wolfe, Rutgers University, Camden, NJ, United States, Srinivasa Puranam

Modern slavery and human trafficking are very prevalent in supply chains. Advancements in technology have yielded new methods and opportunities for identifying and impeding human trafficking. For example, data mining techniques and machine learning algorithms can detect patterns and trends within extensive data sets, making it possible to uncover potential trafficking networks and high-risk areas. This project aims to understand how technology simultaneously facilitates and obstructs human trafficking within a geopolitical understanding.

In focusing on the intersection of geopolitics, the use of technology, and modern slavery, the first objective aims to lay a solid groundwork for the research through a detailed literature review and environmental scan. This crucial step will help pinpoint the existing knowledge, gaps, and potential opportunities in the field. Delving into the business models, supply chain intricacies, and the interplay of political, legal, and regulatory systems highlights the necessity of grasping the wider context where technology and human trafficking converge by looking at human trafficking through the lens of the supply chain – where human trafficking begins, who are the key players, and how it gets to its final consumers.

3 - Adoption of innovative technology in the last-mile delivery of non-profit food rescue organizations

Ahana Malhotra, McMaster University, Hamilton, ON, Canada, Elkafi Hassini

We investigate an innovative direct-to-client last mile delivery for non-profit food rescue organizations that we call Crowdfeeding. Capitalizing on the existing successful ride sharing platforms, a driver collects surplus food from retail outlets and delivers it to food bank clients that are located on the route of their regular customer deliveries. Interviews with employees of food rescue organizations in Canada have been conducted and analyzed. Empirically grounded analytics is used to gain insights and data analysis is conducted by the Technological Acceptance Model (TAM) for analyzing cases. In the second part, empirical analysis highlighting the performance of the implementation of crowdfeeding. Our findings suggest that crowdfeeding is an effective strategy for organizations that care about minimizing their impact on the environment, maximizing the availability of nutritious meals, and decreasing food waste. In addition to our theoretical contributions to the food rescue literature, we offer an empirical account of our theoretical framework, proposals for future research and practical suggestions to managers handling food rescue operations.

Keywords: last-mile delivery, non-profit organization, food rescue, Technological Acceptance Model, case study research

WA24

Summit - 344

Port Operations and Freight Logistics

Contributed Session

Chair: Pirmin Fontaine, KU Eichstaett-Ingolstadt, Ingolstadt School of Management, Ingolstadt, Germany

1 - A Cost Focused Machine Learning framework for replenishment decisions under transportation cost uncertainty

Pirmin Fontaine, KU Eichstaett-Ingolstadt, Ingolstadt School of Management, Ingolstadt, Germany, Daniel Müllerklein, Janosch Ortmann

Companies worldwide seek a balance between risks and cost-efficiency in their supply chains. Due to the increase in extreme weather events, global inland waterway transport disruptions gained growing attention as shipment carriers enforce contractual surcharges to account for capacity losses. To improve efficiency and resilience, one open question is, therefore, when and how much to transport from which of the multiple suppliers considering their lead time differences and account for the transportation cost uncertainty driven by the enforced surcharges.

We formulate this problem as a stochastic Inventory Routing Problem with Direct Deliveries and introducing a new Cost Focused Machine Learning (CFML) framework. Compared to existing approaches, we perform a hyperparameter tuning where the costs of applying the resulting transportation decisions are optimized instead of the prediction score of the individual decisions of the machine learning model.

We evaluate our CFML in a case study, based on a chemical company at the border of the river Rhine. Relevant features for our CFML include the inventory position, historical water level, their trends, and predictions. While traditional learning approaches result in inefficient policies, we show that our CFML can reduce costs by 18% compared to classical machine learning frameworks and more than 20% compared to a standard (s,Q)-reorder policy representing industry standard.

2 - Optimal Capacity Decisions for Competing Forwarders under Transshipment

Jiyong Kim, Seoul National University, Seoul, Korea, Republic of, Byeongkwon Lee, Kunsoo Park, Kwanghun Chung

Carriers and forwarders are major players in the air cargo industry, typically entering into contracts for the carriers' capacity. Forwarders serving the same routes compete for this capacity. At the same time, to better align with demand, they may choose to transship with each other. To better understand these dynamics, we model the forwarder's problem of finding the optimal reservation quantity under capacity substitution and transshipment, along with the optimal transshipment price. Furthermore, we incorporate the carrier's problem and identify a contract that can coordinate the entire supply chain.

3 - Equilibrium Analysis of Shipping Market Priority Competition with Multiple Stakeholders

Xinyue Pu, Tsinghua University, Beijing, China, People's Republic of, Xi Lin, Xiwen Bai

With the rapid growth in trade volume and the expansion of port service regions, port competition has intensified, prompting ports to adopt strategic measures such as berthing priority. This paper investigates the impact of berthing priority on port congestion, stakeholder competition, and social welfare by developing an analytical tri-level game-theoretical model to explore the maritime equilibrium in a port-carrier-cargo-owner system. Specifically, cargo owners, operating under perfect competition, make cargo assignment decisions to minimize their costs. At the carrier level, a Cournot-Nash game is modeled where carriers decide on fleet allocation to balance the revenue from cargo transportation with the costs induced by port congestion. Ports determine the provision of priority to enhance their throughput. Through this intricate game-theoretical framework, we derive market equilibrium using numerical experiments. Our research reveals the potential impact of berthing priority on the entire system, considering carriers' complex decision motivations, and offers decision support from a managerial perspective.

4 - A Stochastic Optimization Model for Tramp Scheduling with Many-to-many Matching and Speed Optimization

Liangqi Cheng, Tsinghua University, Beijing, China, People's Republic of, Xi Lin, Xiwen Bai

The tramp shipping industry involves large volumes of homogeneous cargoes transported around the world. Tramp companies route vessels to pick-up and deliver these cargoes in order to minimize total costs, but the scheduling process becomes increasingly challenging due to the uncertainties and risks in the tramp market. In addition, long transportation time and dispersed pick-up and delivery regions cause delayed responses to market volatilities and exacerbate the complexity of short-term decision-making. This study introduces a tramp scheduling model considering stochastic future revenues to enhance short-term decisions. The model employs “many-to-many” matching in the tramp market, where each vessel can transport multiple cargoes simultaneously, and a single cargo can be split and delivered by multiple vessels. Initially, a two-stage stochastic arc flow formulation is presented within a space-time network, enabling vessel speed optimization during arc selection. Subsequently, Dantzig-Wolfe Decomposition is utilized to manage the extensive set of variables, resulting in a path flow master problem and pricing subproblems, each solved for a specific vessel and scenario. A branch & price solution method is devised, incorporating a label-setting algorithm for the subproblem in each column generation iteration that can determine the cargo quantity decisions brought by split deliveries. This label-setting algorithm is developed through an in-depth analysis of the extreme structure of delivery quantities on a specific route. Extensive numerical experiments are conducted to demonstrate the efficiency of the proposed algorithms. The computational results also uncover the temporary idle and repositioning behaviors of vessels across different tramp markets.

5 - Dynamic Berth Allocation Policies in Deep-Sea Terminals

Orkun Tunay, Rotterdam School of Management, Rotterdam, Netherlands, Pieter van den Berg, Rob Zuidwijk, DEBJIT ROY

While deep-sea vessels play an important role for the revenue streams of container terminals, barges and feeders play an important role in ensuring the functioning of the transshipment and hinterland transportation operations. Besides, the roles of feeders and barges are further leveraged as a cost and emission-efficient alternative to long-haul trucking. These different classes of vessels compete for the same berth space in container terminals. The terminal operators conventionally have the tendency to fully prioritize deep-sea vessels. However, with the increasing incentives and regulatory pressure for the increase of barge and feeder throughput, some terminal operators started to allocate dedicated berths for barges and feeders.
In this paper we will study the berth allocation problem in order to find a pareto-optimality frontier that balances barge throughput and deep-sea vessel waiting times. Barges and feeders which are of smaller volume, occupy a smaller space and have a shorter expected handling time. By using analytical methods from Queuing Theory, we will first demonstrate the pareto-efficiency of different heuristics of berth allocation policies, i.e. dedicated terminal spaces, barge limits, motivated by the current industry practices. Based on the insights that we receive from analytically tractable systems, we will design a Markov Decision Process to propose a better pareto-optimality frontier with dynamic policies.

WA25

Summit - 345

Embracing Uncertainty: Digital Workspaces and Resilience Strategies

Contributed Session

Chair: Siyi Yang, Fudan University, Shanghai, N/A

1 - Visualization of Ideation Process Considering Thinking Styles in Discussion

Mayato Hattori, Waseda University, Tokyo, Japan, Kotomichi Matsuno, Terumasa Mastuyuki, Kenta Nakagawa, Yoshikuni Edagawa, Takahiro Ohno

In contemporary society, where the generation of innovation is an urgent issue, balancing "knowledge exploration" and "knowledge exploitation" are indispensable. In the field of ideation, where ideas are generated, it is necessary to repeatedly cycle through divergent and convergent thinking and discuss objectives and means bidirectionally to produce effective ideas. Although the thought patterns of Causation and Effectuation have been systematized, the application of these methods remains highly subjective, often relying on the experience and intuition of facilitators. Proficient facilitation contributes significantly to the generation of innovative ideas and to the efficiency of discussions. Therefore, there is a need for a reproducible and effective facilitation methodology. This research aims to establish an engineering-based ideation method that focuses on the interaction between divergent and convergent discussions and the interplay between objectives and means. Methodologies are proposed to infer thought patterns and discussion states.

In workshops aimed at creating new businesses in Japan, surveys on thought patterns were conducted before and after ideation, and discussion audio and brainwaves were measured. Bayesian network analysis revealed that thought patterns change due to brain activity. Additionally, using audio data, discussion states were classified into four types: "divergent," "convergent," "goal-driven," and "means-driven," and their relationships with thought patterns were analyzed. This study contributes academically by providing a structured framework for understanding and analyzing ideation processes, and practically by offering tools and methods that can enhance the effectiveness and reproducibility of innovation activities in real-world settings.

2 - Using Disruption Intensity to Measure System Resilience

Daniel Romero, SUNY Polytechnic Institute, Utica, NY, United States, Alex Savachkin, Weimar Ardila, Jose Navarro

Measuring the resilience of a system is essential in evaluating its ability to withstand disruptive events. We present a resilience metric that explicitly incorporates the intensity of a disruptive event to provide a more accurate estimation of system resilience. A comparative analysis between the proposed metric and average performance resilience metrics for linear and nonlinear loss and recovery functions is presented. The new metric is evaluated using a testbed of a community affected by a pandemic influenza outbreak, where nonlinear relationships between resilience capacities and disruption intensity are captured and used for pandemic control.

3 - Overcoming Data Contribution Uncertainty: Relational Contracts for Effective Data Market

Siyi Yang, Fudan University, Shanghai, China, People's Republic of, Jinzhao Wang, Yifan Dou, Lihua Huang

The rapid growth of data-driven industries like Artificial Intelligence has sparked a demand for data circulation and transactions, leading to the emergence of data markets. A major challenge in data market is achieving consensus between buyers and sellers regarding the price of data. This primarily stems from that the value buyers can ultimately derive from the data depends on the collaborative efforts of both parties in the data transactions, which, however, is inherently difficult to measure. Such uncertainty underscores the necessity for developing sophisticated contract forms capable of accommodating the unique characteristics of data transactions. This paper introduces the relational contract as a potential solution to the above challenge, which is a well-established informal contract form that relies on self-enforcement by both parties through repeated interaction. Our analytical results indicate that a transaction is only viable when the reference price is relatively low. Upon a data transaction, the contract form choice depends on the expected data contribution. If it is relatively high, relational contract is optimal; otherwise, only a formal contract can be used. Notably, utilizing relational contracts leads to a higher performance and profits of both the buyer and seller than formal contracts. Thus, to build long-term and efficient trading relationships, market planners can encourage both transaction parties to adopt relational contracts by offering several reasonable reference prices. This not only alleviates the dilemma of pre-selecting the optimal price but also effectively incentivizes both parties to invest in the transaction process and unlock the value of data.

WA26

Summit - 346

Workplace Behavior and Organizational Dynamics

Contributed Session

Chair: Yuan Li, University of Tennessee, Knoxville, TN, United States

1 - Investigating Impact of Personality Traits, Workplace Ostracism and Perceived Organizational Justice on Work Procrastination Among Professors: Mediating Role of Cyberloafing

Marwah Rafiq, University of science and technology beijing, beijing, China, People's Republic of, Jianhua Yang, Sahafiul Bashar

This study investigates the effects of workplace ostracism, perceived organizational justice, and personality factors on workplace procrastination. This study additionally investigates into the mediating function of cyberloafing in these interactions. This study examines at the relationships between workplace ostracism, personality traits, perceived organizational justice, cyberloafing, and work procrastination. When conducting the empirical component, this study used the partial least squares structural equation modeling approach using a sample of 251 academics from Pakistani and German institutions. The findings show that personality traits have a significant impact on employee cyber-loafing. Employees who face workplace ostracism and poor perceived organizational justice are less productive and more likely to procrastinate at work. Furthermore, this study provides useful information for investigating cyberloafing's mediation function in the relationship between these variables and work procrastination. Likewise, when interacting with educational institutes, it was found that professors might be treated unfair at times and receive less support from their supervisors which leads them to face workplace ostracism and perceived organizational justice, causing deviant behavior in them. This study adds to the “Theory of Organizational justice” by working on perceived organizational justice faced by professors. Furthermore, this research adds to the “theory of interpersonal behavior” and the “theory of planned behavior” by exploring the relationships of workplace ostracism and personality traits with cyberloafing and eventually work procrastination.

2 - A recommender system for employee retention based on a new feature selection strategy

Gulser Koksal, TED University, Ankara, Turkey, Nagihan Taskiran

This study addresses the critical issue of employee retention in organizations, emphasizing the importance of personalized strategies over generic approaches. Traditional retention strategies may not effectively cater to individual employee needs and priorities. To tackle this, the study proposes an enhanced recommendation system utilizing a combination of clustering and ensemble learning techniques. By identifying "precious" employees through Fuzzy C-Means clustering, the study selects the most influential attributes affecting churn using a novel feature selection method. Various classification algorithms are then employed to predict potential churners, including Random Forest and Gradient Boosting. Moreover, a unique recommendation approach is introduced, comparing similar non-churner and potential churner profiles to suggest personalized retention strategies. The study ultimately provides HR departments with a set of attributes tailored to each potential churner employee, categorized by experience level, aiding in proactive retention efforts.

3 - Negative Feedback in Corporate Communication: a Social Sensitivity Perspective

Yang Jiao, ShanghaiTech University, Shanghai, China, People's Republic of, Ruijie Zhang, Lifeng Yang

Effective communication within corporate management is crucial for operational excellence. Often, project managers may overlook existing issues that are apparent to their subordinates, leading to dissatisfaction and negative feedback from employees. This discrepancy can result in substantial losses for the firm if not addressed. The reluctance of employees to provide high-quality feedback can be attributed to psychological barriers related to social sensitivity, a factor not extensively explored in operations management literature. This study introduces a behavioral perspective that incorporates social sensitivity into the analysis of corporate communication dynamics. We employ stylized models under various scenarios to simulate the decision-making processes of employees when considering (1) whether to provide negative feedback to project managers, and (2) the extent of the provision. Initially, a simplified model shows the basic relationship between project completion and firm returns. Incorporating social sensitivity reveals its importance in enhancing feedback quality and optimizing effort, thus improving profitability. Additionally, we finds that cooperation among employees leads to higher feedback quality and project returns by leveraging diverse perspectives, and different social sensitivity degrees affect employees' contribution. This paper concludes by highlighting the need for effectively leveraging social sensitivity and promoting cooperation to enhance project performance and profitability, offering practical insights for managerial decision-making.

4 - Employees’ affects and beliefs as antecedents of OCB towards change implementation

Bernardo Quiroga, West Virginia University, Chambers College of Business, Morgantown, WV, United States, Andrés Raineri Bernaín

This paper uses an attitude theory framework to analyze the simultaneous contribution of employees’ Change Readiness and Change Cynicism beliefs, as well as their positive and negative affects towards change, as antecedents of employees’ Organizational Citizenship Behaviors directed Towards Change (OCBTC) implementation. Using a time-lagged methodology, this study examines, in a sample of 402 employees, the additional contribution that different cognitive and emotional components of employees’ attitudes towards change have in explaining employees’ OCBTC. Relative Weight Analysis results indicate that core affects explain variance of OCBTC beyond that explained by Change Readiness and Change Cynicism. As well, Change Readiness and Change Cynicism beliefs do add to each other’s contribution in predicting the criterion variables. These results support the use of an attitude theoretical framework for researchers to study employee reactions to change, where both cognitive and affective components are considered as antecedents of employee behaviors. Furthermore, results suggest that managers should foster among their employees' both the cognitive and affective components of appropriate attitudes towards change, to help them engage in behaviors aiming to facilitate organizational change implementation.

5 - The Impact of Psychological Need Satisfaction and Frustration on Security-related Stress and ISP Violations

Yuan Li, University of Tennessee, Knoxville, TN, United States

Security-related stress (SRS) causes employees' violations of information security policies (ISPs). Its antecedents have not been adequately analyzed, as past research assumed its existence but focused on the coping responses to SRS. We suggest that the causes of SRS are more complex, as employees' perceptions about the motivations and basic psychological needs play vital roles in determining SRS.

Building on the recent literature on psychological need satisfaction and frustration, we develop a model to examine the antecedents to SRS. We argue that the intrinsic and extrinsic motivation factors, such as perceived legitimacy of ISP, perceived value congruence, perceived punishment, and perceived reward, influence psychological need satisfaction and frustration. Psychological need satisfaction and frustration then influence SRS, which in turn affects ISP violations.

A survey on 271 employees in the United States was conducted to empirically test the research model. The results show that perceived legitimacy and perceived value congruence increase need satisfaction and decrease need frustration. While punishment decreases need satisfaction, it does not affect need frustration. Reward increases both need satisfaction and frustration, exhibiting a punitive side effect. On the other hand, need frustration is a primary cause of SRS whereas the effect of need satisfaction on SRS is partialed out. As expected, SRS increases ISP violations.

This study highlights the mediating roles of psychological need frustration in determining SRS and ISP violations. Due to the difference between need frustration and satisfaction, the study emphasizes the value of analyzing the causes and consequences of need frustration in ISP requirements.

WA27

Summit - 347

Innovative Mechanisms

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Shivam Gupta, University of Nebraska Lincoln, Lincoln, NE, United States

Co-Chair: Like Bu, The University of Texas at Dallas, Richardson, TX, United States

1 - Service Systems with on-Demand and Reserved Servers

Zhichao Feng, Hong Kong Polytechnic University, Hung Hom, Hong Kong, Milind Dawande, Ganesh Janakiraman, Anyan Qi

Many modern queueing-based service systems have a choice of using either long-term reserved capacity or short-term on-demand capacity; examples include the use of cloud-computing resources by firms, healthcare professionals by post-acute-care services, and customer-service agents by call centers. We study a canonical M/M/s queueing system that employs both reserved and on-demand servers – the number of reserved servers is decided at the beginning of the time horizon while the number of on-demand servers is decided dynamically as needed, in real time. We study two problems: minimizing the infinite-horizon discounted cost incurred in the hiring of servers and in the waiting of the jobs in the queue, and minimizing the long-run average cost. Using a sample-path analysis and a vanishing-discount approach, we show that the optimal on-demand capacity control in both problems is a threshold-based bang-bang policy: If the number of jobs in the system is below a threshold, then no on-demand servers are employed. Otherwise, the number of on-demand servers is chosen such that no jobs wait (i.e., the number of servers in use equals the number of jobs). Significantly, we obtain a closed-form expression for the cost of any such policy for a given number of reserved servers and derive an algorithm for obtaining the optimal number of reserved servers and the threshold. Our analysis provides practical rules-of-thumb for managing capacity and waiting costs in queueing systems that have access to both long-term and short-term server capacity.

2 - Combating Overbilling in Outsourced Projects: A Dynamic Auditing Mechanism

Like Bu, The University of Texas at Dallas, Richardson, TX, United States, Shivam Gupta, Milind Dawande, Ganesh Janakiraman

The practice of overbilling, where an external agent charges a firm more than what is deemed reasonable for the outsourced service, has been well-acknowledged as a notorious and long-standing problem across major industries. We consider a repeated-interaction principal-agent setting and develop a non-monetary dynamic auditing mechanism to mitigate overbilling.

3 - Pay with Your Data: Optimal Data-Sharing Mechanisms for AI Services

Seetharama Chandrasekhar Manchiraju, Eli Broad College of Business, Michigan State University, East Lansing, MI, United States, Sameer Mehta, Milind Dawande, Ganesh Janakiraman

Rapid advances in Machine Learning (ML) have led to a proliferation of Artificial Intelligence~(AI) services offered by firms. To develop a valuable AI service, a firm must build an accurate ML model which, in turn, requires a large amount of training data. Present-day firms usually obtain this data by offering incentives to consumers to share their data during the initial development phase of the AI service, and then use that data to re-train the ML models to improve the quality of the service. Consumers, on the other hand, incur privacy costs for sharing their data. Inspired by AI services such as speech-to-text conversion offered by Google, and ChatGPT and DALL-E offered by OpenAI, we analyze two popular data-sharing mechanisms that firms employ in practice: manual data-sharing and algorithmic data-sharing. In the former approach, consumers decide the amount of data to share with the firm, whereas in the latter, the firm uses algorithmic data-redaction — an established approach used by technology firms such as Amazon, IBM, and Oracle — to identify and censor sensitive segments of data, and determine the amount of data collected from consumers. For both the data-sharing approaches, we obtain revenue-maximizing mechanisms for the firm and analyze the fundamental differences between the two approaches in terms of the revenue accrued by the firm, the consumer surplus, and the volume of data collected. Our analysis uncovers several interesting economic effects.

4 - Combating Overbilling in Outsourced Projects: A Dynamic Auditing Mechanism

The practice of overbilling, where an external agent charges a firm more than what is deemed reasonable for the outsourced service, has been well-acknowledged as a notorious and long-standing problem across major industries. We consider a repeated-interaction principal-agent setting and develop a non-monetary dynamic auditing mechanism to mitigate overbilling.

WA28

Summit - 348

Innovative Operations for Social Goods and Sustainability

Contributed Session

Chair: Qianqian WANG, N/A

1 - is Innovation Competition vs. Innovation Cooperation Strategic Choice of Platform Considering Complementors Innovation

Qianqian WANG, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Sai-Ho Chung, Qiang Chen

The level of innovation among complementors, i.e., third-party firms providing products, technologies, or services on digital platforms, plays a crucial role in sustaining the competitive advantage of these platforms. To encourage complementors to innovate, digital platforms have employed either innovation competition or innovation cooperation strategies. However, the effectiveness of these strategies in incentivizing complementors to innovate remains a topic of extensive debate, leading to practical ambiguity between complementors and digital platform firms. To address this issue and compare the innovation incentive effects of the two strategies, we develop a Stackelberg game model. Considering innovation spillover effects and innovation synergy costs, we investigate the innovation decisions and benefits of strong and weak complementors under the two digital platform strategies. Our preliminary findings suggest that strong and weak complementors respond differently to the innovation incentive strategies of digital platforms. The innovation competition strategy provides better incentives for weak complementors to innovate, while the innovation cooperation strategy is more effective in incentivizing innovation among strong complementors. We observe that higher competition intensity leads to increased innovation efforts by complementors, but the innovation spillover rate hampers their innovation incentives. Digital platforms sharing complementors' innovation costs can effectively enhance their motivation to innovate, but co-innovation with complementors does not necessarily increase their innovation efforts, as it is influenced by co-innovation costs. These findings offer valuable insights for complementors in determining their innovation efforts and for digital platform firms in optimizing their innovation incentive strategies.

2 - Partial Vertical Ownership, Specific Investment, and Control in a Platform Supply Chain

Chenchen Hua, School of Economics and Management, Southeast University, Nanjing, China, People's Republic of, Lindu Zhao

The acceleration of digital and intelligent technology development and the growth of customers' individual needs promote transformation and upgrading in manufacturing industry. Researchers have devoted great efforts to working on centralized and decentralized supply chains, making contributions to operation and management area. However, much less attention has been focused on supply chains with financial interconnectedness, such as partial vertical ownership, which is also a common business phenomenon between firms in manufacturing area.

This paper studies the impact of partial vertical ownership (PVO) in a supply chain consisting of a upstream manufacturer and a downstream industrial internet platform. The industrial internet platform holds non-controlling partial vertical interests of the manufacturer in order to establish a specific relationship between them. The platform is able to acquire superior information about the future market demand and place an order with the manufacturer who is capable of fulfilling the order. Passive partial vertical ownership may encourage the manufacturer to make specific investments for the satisfaction of diversified requirements in the market. We work on the impact of PVO on manufacturer's specific investment level and the platform's operational cost reduction level with game theory model.

This research may possibly contribute to the theory and practice of firms with financial interconnectedness in a supply chain to some extent, though limitation remains. Further studies are needed to broaden the applicable scope of this research, and to figure out the influence of PVO in a supply chain in more complicated scenarios.

3 - Capacity sharing strategies for competing manufacturers: component-sharing or technology-sharing?

Xuzheng Li, School of Economics and Management, Southeast University, Nanjing, China, People's Republic of, Lindu Zhao

Shared manufacturing is an important part of service-oriented manufacturing and introduces an innovative business model for allocating manufacturing resources in the sharing economy era. This business model includes a component-sharing mode, which uses shared production lines, and a technology-sharing mode, which involves sharing product design capabilities. Within this framework, manufacturers with different capacity can not only design and produce a complete product family independently, but also cooperate with competitors to produce some products in a product family. Specifically, manufacturers with limited capacity can produce non-core components independently, purchase core components at a lower cost from leading manufacturers through a component-sharing mode, and subsequently assemble the product with additional processing. Alternatively, through the technology-sharing mode, manufacturers with limited capacity can acquire technologies for core components at a reduced cost by paying fixed license fees and royalties to leading manufacturers. They can then innovate and enhance these technologies to develop new products. This study adopts an oligopolistic competition modeling framework to construct game-theoretic models of manufacturers under different modes and derive the corresponding equilibrium solutions. Based on profit analysis, we identify the optimal mode selection strategies for manufacturers with different capacity. This research broadens the scope of studies in shared manufacturing and has significant implications for the operational decisions of manufacturers.

4 - Pricing and Operations Planning for New Product Launch: Introducing Adopters' Variant Willingness to Pay

Alejandro Lamas, NEOMA Business School, Mont-Saint-Aignan, France, Sara Jahanmir

For sustainable profit growth and competitiveness, firms dedicate significant resources to development and launch of new technologies. Setting the right price for these innovations is an enormous challenge, which requires (i) a clear understanding of the social interaction influencing adopters' purchasing decisions, and (ii) coordinating marketing and operations so that products are available and affordable for adopters, especially for those who are more resistant to higher prices as the innovation moves through its life cycle. Building on this, we study the joint pricing and operations planning of firms when launching a new product in a market where the adopters' willingness to pay decreases over time. We combine and extend the traditional Bass model for diffusion and lot-sizing problems to price-sensitive demand settings. As the resulting formulation suffers from non-linearity, we propose an efficient linear approximations-based methodology for price/diffusion relationships. We show that both optimistic and pessimistic biases due to ignoring adopters' heterogeneous willingness to pay can result in losses of more than 20% relative to our modeling. We also show that ignoring such heterogeneity can blur the financial returns from launching new products and thus affect the commercial success of innovations.

5 - Coordination of Product-Service Supply Chain in a Framework of Profound Integration of Manufacturing and Service Industries

Yanling Feng, Beijing university of posts and telecommunications, Beijing, China, People's Republic of

The continuous evolution of the integration between manufacturing and service industries is paving the way for new opportunities for economic development. The quality of service provided by sales service integrator has become a critical determinant of product demand in the ensuing product-service supply chain. However, there exists an information asymmetry concerning service levels among the parties involved in the supply chain. Since the integrator has direct access to the consumer market, it possesses the informational advantage regarding the actual service level it provides.

This study examines a two-echelon supply chain composed of a manufacturer and a sales service integrator. The service cost sharing contract is developed with the understanding that integrator is under a lot of pressure to cover the service cost alone, which results in a lack of motivation to raise the service level. The integration of blockchain technology, with its inherent information traceability, into the product-service supply chain management is a strategic move aimed at mitigating service level information asymmetry. This approach ensures the seamless execution of contracts, even when information is incomplete.

The results reveal that service cost-sharing contracts can serve as effective incentives for integrator to augment service quality, adjust product pricing, and achieve supply chain coordination and optimization. It provides theoretical proof that blockchain technology can effectively alleviate the issue of information asymmetry, thereby boosting the overall profit margin of the entire supply chain system.

WA29

Summit - 420

Computational Integer Optimization

Invited Session

OPT: Integer and Discrete Optimization

Chair: Daniel Bienstock, Columbia University, New York, NY, United States

Co-Chair: Daniel Bienstock, Columbia University, New York, NY, United States

1 - Polynomial TIME Algorithms for Mixed-Integer Inference Problems

Andres Gomez, University of Southern California, Los Angeles, CA, United States, Shaoning Han

We discuss classes of nonlinear mixed-integer optimization problems arising in statistical inference that admit polynomial time algorithms. The problems in question include outlier detection, sparse inference and data imputation, and have in common an underlying submodular structure. We develop polynomial time algorithms relying on a parametric algorithm to solve related convex optimization problems and exploiting de Lovasz extension of the underlying submodular functions.

2 - A Two-Stage Stochastic Integer Programming Approach for Open-Pit Mine Scheduling Under Geological Uncertainty

Patricio Lamas, AlicantoLabs, La Reina, Chile, Marcos Goycoolea, Eduardo Moreno

In open-pit mine scheduling, it is assumed that a mine corresponds to a discrete set of material blocks. Each block is characterized by a set of attributes such as weight, volume, grade of mineral, etc. Thus, the scheduling problem consists of answering three questions: WHICH blocks to extract; WHEH to extract those blocks; and WHERE to send the extracted blocks. The goal is to maximize the net present value of the mine project. Traditionally, the open-pit mine scheduling problem has been solved assuming that the attributes of each block are perfectly known before the project starts. Such an assumption does not necessarily hold, given that, in practice, the blocks attributes are only approximately known through geostatistical methods using drill-hole information.

We propose a two-stage stochastic programming approach for solving the problem under geological uncertainty. In a first stage, the WHICH and WHEN questions are answered, while, in a second stage, when the uncertain geological attributes are realized, the WHERE question is answered.

We show that our proposed two-stage stochastic programming problem can be efficiently solved with the Bienstock-Zuckerberg decomposition algorithm. This decomposition algorithm has been extensively used for solving traditional, deterministic open-pit mine scheduling problems.

Moreover, we provide Mak et al. Monte Carlo bounds for determining the quality of the solution obtained by solving our two-stage stochastic programming problem.

Our computational results suggest that our approach provides schedules with significantly larger expected net present values than those provided by the traditional deterministic approaches.

3 - Cardinality-Constrained Convex Quadratic Programming Problems Under Sparsity Conditions

Tongtong Chen, Columbia University, New York, NY, United States

We present results on cardinality-constrained convex quadratic programming problems under sparsity conditions. In the problems, we consider continuous variables partitioned into a family of subsets; only a constrained count of such subsets are allowed in the feasible solutions. Motivated by problems from two sources, (1) an instance in QPLIB and (2) the cardinality-constrained portfolio optimization problem, we introduce a new metric for sparsity, namely, the treewidth of the intersection graph for the transpose of our constraint matrix. We provide complexity results and approximation algorithms for those types of problems.

4 - Parallelized Conflict Graph Cut Generation

Yongzheng Dai, The Ohio State University, Columbus, OH, United States, Chen Chen

A conflict graph represents logical relations between binary variables, and effective use of the graph can significantly accelerate branch-and-cut solvers for mixed-integer programming (MIP). In this paper we develop efficient parallel conflict graph management: conflict detection; maximal clique generation; clique extension; and clique merging. We leverage parallel computing in order to intensify computational effort on the conflict graph, thereby generating a much larger pool of cutting planes than what can be practically achieved in serial. Computational experiments demonstrate that the expanded pool of cuts enabled by parallel computing lead to substantial reductions in total MIP solve time, especially for more challenging cases.

5 - Incorporating Service Reliability in Multi-depot Vehicle Scheduling: A Chance-Constrained Approach

Merve Bodur, University of Edinburgh, Edinburgh, United Kingdom

The multi-depot vehicle scheduling problem (MDVSP) is a principal planning challenge for transit agencies. We introduce a novel approach to MDVSP by incorporating service reliability through chance-constrained programming (CCP), targeting the pivotal issue of travel time uncertainty and its impact on transit service quality. We propose an exact branch-and-cut scheme to solve our CCP model. We present several cut-generation procedures that exploit the underlying problem structure and analyze the relationship between the obtained cut families. Additionally, we design a Lagrangian-based heuristic to handle large-scale instances reflective of real-world transit operations. Our empirical evaluation demonstrates the superiority of our stochastic variant as well as the computational benefits of our methodologies.

WA30

Summit - 421

Computational Nonlinear Programming

Invited Session

OPT: Computational Optimization and Software

Chair: Hande Benson, Drexel University, Philadelphia, PA, United States

1 - A quantum-inspired quadratic programming solver with large-scale GPU implementations

Yuxiang Peng, University of Maryland, College Park, MD, United States, Jinglei Cheng, Samuel Kushnir, Pengyu Liu, Jiaqi Leng, Hanrui Wang, Xiaodi Wu

Quantum and quantum-inspired algorithms have demonstrated significant advantages for numerous non-convex optimization problems. Compared with conventional quantum algorithms, quantum-inspired algorithms do not require access to quantum computers yet still exhibit advantageous performance against classical methods. Specifically, quantum-inspired algorithms offer a more flexible design space, enabling the use of cutting-edge computational hardware like GPUs. In this talk, we introduce a new solver based on a recently proposed quantum-inspired algorithm for quadratic programming. Our implementation is optimized for large-scale GPU deployment, providing substantial computational speedups at low costs. We benchmark our solvers on both public and synthetic datasets and demonstrate that our solutions are superior in runtime and solution quality compared to other solvers on large problem instances.

2 - Multi-Objective Optimization-Based Anonymization of Structured Healthcare Data for Machine Learning Applications

Yusi Wei, Drexel University, Philadelphia, PA, United States

Healthcare data plays a crucial role in machine learning (ML) for clinical research; however, the risk of multiple types of privacy attacks still remains when sharing such data. Techniques have been proposed to further protect data privacy, but there is still a need for models that can simultaneously achieve protection against attacks while maintaining the information value of healthcare data to ensure the usefulness of ML models. In this paper, we propose and solve a multi-objective optimization model for minimizing information loss and enhancing protection against attacks on healthcare data for the diagnosis of sepsis. We compare our model with existing algorithms to assess information loss and the number of patients subject to linkage or homogeneity attacks. The results indicate that our proposed model achieves the lowest information loss and effectively mitigates the risk of attacks by reducing more number of patients susceptible to these attacks in healthcare data compared to the alternative algorithms. We also leverage our anonymized quasi-identifiers along with other variables present in healthcare data for the prediction of sepsis. The findings demonstrate that our model does not adversely affect the prediction of sepsis using ML models when utilizing small values of k for k-anonymity, such as k=5, 10, 15, and 20. This study affirms that our model provides a viable solution for maintaining patient privacy while still allowing for effective ML prediction. Future studies can be built upon our work to develop models that can preserve patient privacy while maintaining the high performance of ML models.

WA31

Summit - 422

Data-Driven Robust Optimization Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Zhi Chen, The Chinese University of Hong Kong, Shatin, Hong Kong

1 - Prescriptive Analytics for Queue Optimization - An Optimization-Based Paradigm for Modelling Queues

Peng Wang, School of Business, Singapore University of Social Sciences, Singapore, Singapore, Chaithanya Bandi, Gar Goei Loke, Taozeng Zhu

Queueing networks occur in many contexts, leading to the need to control and optimize aspects of the network, such as routing of jobs and capacity control. Queueing theory, as it stands, focuses on the analysis of the stochastic properties of a network, leveraging upon them to perform optimization. These methods of analysis, bespoke for the network, rapidly increase in theoretical complexity with network size, making the researchers indispensable in the process. These factors impede the wider use of its techniques in the automation-driven business world. In this paper, we present a fundamentally novel approach to queueing, with the goal of creating a unified framework to solve optimization problems in queueing networks for a large class of queueing problems, that would lend itself to a tractable package that is accessible to business users, untrained in queueing theory. We do so by proposing a novel set of primitives for modelling queues, founded on optimization principles, leaving the correlated stochastics of queues to be resolved as a constrained problem. We also introduce a new state variable, which we call as "present delay" that tracks waiting time at each node. This allows much of the dynamics to decompose into a linear form, which improves tractability. Finally, we handle the stochastics by leveraging emergent techniques in robust optimization that achieves tractable solutions in stochastic optimization formulations. This also hints at good out-of-sample performance. Our final model is one that can solve any connected network of general queues in polynomial time.

2 - Distributionally Robust Optimal Open-Loop Policy in Backlog Inventory Models with Stochastic Lead Times

Jingjun Men, Chinese University of Hong Kong, Hong Kong, Hong Kong, Xin Chen, Daniel Zhuoyu Long

In this paper, we address an inventory management problem characterized by random demand and lead times. Recognizing the inherent uncertainty and potential ambiguity in their distributions, we assume only the marginal distributions of demand and lead times are known. Our objective is to minimize the worst-case expected cost under these conditions, a challenge compounded by uncertain lead times. We approach this by optimizing across all open-loop control policies. Initially, we demonstrate that our inventory model can be formulated as a Mixed Integer Linear Programming (MILP) problem. To streamline the process further, we establish a necessary and sufficient condition for 'no order crossover' in the worst-case distribution. This condition allows the problem to be transformed into a shortest path problem on a directed acyclic network, which can be solved efficiently. We discover that the optimal open-loop policy can be effectively achieved by considering the order quantities as the cumulative sum of consecutive incoming orders over several days. Numerical examples are provided to demonstrate the practical effectiveness of our approach.

3 - Service Oriented Considerate Routing: Data, Predictions and Robust Decisions

yue zhao, Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Zhixing Luo, Stanley Lim, Caihua Chen, Melvyn Sim

This paper focuses on improving service oriented routing by addressing the nuanced challenge of punctuality through the consideration of couriers' ability to ensure on-time deliveries. We utilize a comprehensive real-world dataset from a cold chain logistics firm for analysis. Our empirical investigation indicates that relying solely on travel distance is inadequate for accurate delivery time prediction. We highlight critical elements, including couriers' fixed effects and workload, as key covariates to improve prediction performance. Distinguishing our work from existing literature, we integrate couriers' workload and location familiarity into our Service Oriented Routing model to enhance predictions of delivery times. We introduce the Courier Assigned Location Mismatch (CALM) metric as a less intrusive approach to incorporating couriers' location familiarity into their delivery efficiency. We propose the novel Service Oriented Considerate Routing (SOCR) model; by minimizing the CALM metric, couriers are assigned routes within familiar territories to the extent possible within the total routing distance constraint. The considerate routing strategies could potentially reduce the stress couriers face when delivering in unfamiliar areas. Additionally, we develop the connection of the SOCR model with a robust satisficing approach. This strategy guarantees timely deliveries by effectively mitigating the effects of predictive inaccuracies and potential model misspecifications. To solve the SOCR model, we apply Benders decomposition for an exact solution and Tabu Search for a heuristic approach, demonstrating their effectiveness and superior out-of-sample performance. Notably, our heuristic solutions significantly outperform exact solutions of classical vehicle routing problems with deadlines, resulting in substantial improvements in timely delivery performance.

4 - Optimize-via-Estimate: Realizing out-of-sample optimality in data-driven optimization

Gar Goei Loke, Durham University Business School, Durham, United Kingdom, Taozeng Zhu, Ruiting Zuo

We propose a novel methodology that achieves almost zero regret in data-driven optimization, wherein the uncertainty lies in a parametric family, when one restricts to generalizing over a given region of possible truths. We prove that there are no solutions that are unconditionally optimal over all possible true parameters. When restricted over a neighbourhood of parameters, we prove sufficient conditions for optimality - functions of the sufficient statistic of the parameter. We present an optimization problem that yields such an optimal solution, which can be solved efficiently. We also state sufficient conditions for the existence and uniqueness of this solution. Next, we illustrate specificity-sensitivity trade-off in terms of the size of the neighbourhood considered on the specific example of a one-dimension parameter. Finally, we illustrate our model on the newsvendor model to find strong performance when compared against alternatives in the literature.

WA32

Summit - 423

Recent Advances in First-Order Methods and ADMM.

Invited Session

OPT: Global Optimization

Chair: Leandro Maia, Texas A&M, College Station, TX, United States

Co-Chair: Leandro Maia, Texas A&M University, College Station, TX, United States

1 - An Adaptive Block Proximal ADMM for Weakly Convex, Linearly-Constrained Composite Functions

Leandro Maia, Texas A&M University, College Station, TX, United States, David Gutman, Renato Monteiro, Gilson Silva

This work presents an adaptive, cyclic block proximal version of the alternating direction method of multipliers (AP-ADMM) for solving linearly-constrained, weakly convex, composite optimization problems. This method is adaptive to all problem parameters including smoothness and weak convexity constants. We assume that the smooth component of the objective is weakly convex and possibly nonseparable, while the non-smooth component is convex, block-separable, and has easily computable block proximal maps. Each iteration of AP-ADMM consists of: (1) sequentially solving the sequence of all block proximal subproblems, and (2) performing full Lagrange multiplier and/or penalty parameter updates based on a novel update criterion. Without any rank assumptions on the constraint matrices, it is shown that AP-ADMM obtains an approximate first-order stationary point of the constrained problem in $\boldsymbol{O}(\varepsilon^{-3})$ iterations.

2 - Efficiency of a restarted parameter-free FISTA method for strongly convex/convex optimization with provable complexity guarantees

Arnesh Sujanani, University of Waterloo, Waterloo, ON, Canada, Renato Monteiro

This talk presents an efficient restarted parameter-free FISTA (RPF-FISTA) method for solving strongly convex and convex composite optimization problems. RPF-FISTA adapts to the local geometry underlying the problem and checks novel conditions to determine when to restart. The complexity of the method is established and its computational efficiency is demonstrated on numerous vector and matrix problems.

3 - Last iterate convergence for convex function constrained optimization problem.

Digvijay Boob, Southern Methodist University, Dallas, TX, United States, Mohammad Khalafi

In recent times, various works have explored algorithms for convex function-constrained optimization problem. Most of these works provide convergence guarantees for the average-iterate. Such solutions degrade various desirable properties of the algorithmic output such as sparsity requirements or differential privacy guarantees. Moreover, it is common practice to output the last-iterate. Even then, last-iterate convergence guarantees are uncommon in function-constrained optimization problems. In this talk, we present primal-dual type algorithms that converge for nonsmooth and stochastic convex function-constrained optimization problems in the last iterate. In particular, our algorithms do not need to search for the Lagrange multiplier, nor do we need explicit descent in some potential function to get a bound on the dual iterate. Our designs are simple to implement and the gradient complexity of our proposed methods matches the existing state-of-the-art for the average iterate in both nonsmooth and stochastic cases.

4 - A Riemannian ADMM

Jiaxiang Li, University of Minnesota, Minneapolis, MN, United States, Shiqian Ma, Tejes Srivastava

We consider a class of Riemannian optimization problems where the objective is the sum of a smooth function and a nonsmooth function, considered in the ambient space. This class of problems finds important applications in machine learning and statistics such as the sparse principal component analysis, sparse spectral clustering, and orthogonal dictionary learning. We propose a Riemannian alternating direction method of multipliers (ADMM) to solve this class of problems. Our algorithm adopts easily computable steps in each iteration. The iteration complexity of the proposed algorithm for obtaining an ϵ-stationary point is analyzed under mild assumptions. Existing ADMM for solving nonconvex problems either does not allow nonconvex constraint set, or does not allow nonsmooth objective function. In contrast, our complexity result is established for problems with simultaneous nonsmooth objective and manifold constraint. Numerical experiments are conducted to demonstrate the advantage of the proposed method.

WA33

Summit - 424

Learning and Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Qiran Dong, University of California, Berkeley, Berkeley, CA, 94720, United States

Co-Chair: Paul Grigas, UC Berkeley, Berkeley, CA, United States

1 - Beyond Discretization: Learning the Optimal Solution Path

Qiran Dong, University of California, Berkeley, Berkeley, CA, United States, Paul Grigas, Vishal Gupta

Many machine learning methods require minimizing a family of parametrized optimization problems across some hyperparameter space (e.g. weighted binary classification and fair learning). Traditional approaches discretize this space into a grid and solve problems independently, which does not necessarily exploit the solution path structure. We propose a novel and flexible first-order approach, which prescribes a set of basis functions and uses stochastic gradient descent (SGD) with randomized hyperparameters to solve for the *entire* solution path in a single optimization problem. Our method accommodates a variety of basis functions and offers substantial complexity improvements over traditional methods. Specifically, when the solution path is fully interpolable by the prescribed basis, our method requires O(log(1/ε)) and O(1/ε) gradient calls for exact and noisy gradient oracle models respectively, compared to the best-known grid search schemes which require ~O(ε-1/2) and O(ε-3/2) gradient calls. We demonstrate our method's effectiveness on a weighted logistic regression problem with an imbalanced dataset.

2 - Fusing learning and optimization for sustainable power system analytics

Linwei Sang, Tsinghua University, Shenzhen, China, People's Republic of

3 - Route `em and Count `em: A Two-Stage Stochastic Programming Model in Undersea Warfare

Vanessa Sawkmie, University of Wisconsin-Madison, Madison, WI, United States, Jeff Linderoth

The detection of targets in undersea warfare requires successful detection by an active search asset. Maximizing detection likelihood requires strategic placement and routing of the search assets, each of which has unique search capabilities, in the decision theatre over the planning horizon. We develop a two-stage integer stochastic programming model designed to maximize the expected number of targets detected, considering both the stochastic nature of target movements and the probabilistic detection capabilities of the assets. Leveraging the problem structure, we develop a closed-form solution for the second stage of the stochastic program, which allows for large-scale instances of the model to be solved efficiently via Benders decomposition. We conduct computational experiments to demonstrate the efficacy of the approach.

4 - On Solution Structure and Stability of Optimality-Based Clustering

Soroush Akbarijokar, University of Pittsburgh, Pittsburgh, PA, United States, Taewoo Lee

We study a new clustering problem where data points are decisions made by solving some optimization problem and are clustered based on their latent decision-making preferences. We model this problem as a large-size mixed-integer program, which we then reformulate as a more efficient combinatorial problem. We then propose a stability-driven variant of the clustering model to mitigate the impact of data noise and uncertainty on the clusterings. We discuss practical implications of this problem in healthcare applications including eliciting group diet preferences and clustering clinical pathways based on underlying clinical preferences.

WA34

Summit - 425

Lagrangian Methods and Optimization Techniques

Contributed Session

Chair: David Gay, AMPL Optimization, Inc., 900 Sierra Place SE, Albuquerque, NM, 87108-3379, United States

1 - Evaluating Hessians in Parallel

David Gay, AMPL Optimization, Inc., Albuquerque, NM, United States

When solving nonlinear optimization problems, sometimes it is helpful to use the Hessian matrix (of second partial derivatives of the Lagrangian function). Many nonlinear problems exhibit partially separable structure, which can be found automatically, as in the AMPL/solver interface library. This structure invites parallel evaluation. This talk reviews Hessian structure and discusses exploiting it in parallel.

I may not get to Seattle until late Sunday. Please schedule this talk for Monday or later.

2 - Exact Augmented Lagrangian Duality for Mixed Integer Convex Optimization

Avinash Bhardwaj, Indian Institute of Technology Bombay, Mumbai, India, Vishnu Narayanan, Abhishek Pathapati

Augmented Lagrangian dual augments the classical Lagrangian dual with a nonnegative nonlinear penalty function of the violation of the relaxed/dualized constraints in order to reduce the duality gap. We investigate the cases in which mixed integer convex optimization problems have an exact penalty representation using sharp augmenting functions (norms as augmenting penalty functions). We present a generalizable constructive proof technique for proving existence of exact penalty representations for mixed integer convex programs under specific conditions using the associated value functions. This generalizes the recent results for mixed integer linear programming [M. J. Feizollahi, S. Ahmed, and A. Sun, *Math. Program.*, 161 (2017), pp. 365–387] and mixed integer quadratic progamming [X. Gu, S. Ahmed, and S. S. Dey, *SIAM J. Optim.*, 30 (2020), pp. 781–797] while also providing an alternative proof for the aforementioned along with quantification of the finite penalty parameter in these cases.

WA35

Summit - 427

Modeling and Techno-economic Analysis for Vehicle Grid Integration and Vehicle-to-Everything (V2X)

Invited Session

TSL: Intelligent Transportation Systems

Chair: Zhaocai Liu, National Renewable Energy Laboratory, Golden, CO, United States

Co-Chair: Yili Tang, University of California, Berkeley, Hong Kong, ON

1 - Valuing V2x for Customer and Grid Resilience

Zhaocai Liu, National Renewable Energy Laboratory, Golden, CO, United States

Vehicle grid integration (VGI) will bring disruptive and revolutionary impact on grid reliability, efficiency, and resilience. EVs with bidirectional charging capability can be used to power a home and feed energy back into the electricity grid, and thus are able to provide support to enhance system resilience. This presentation will show our modeling and analysis for quantifying the potential resilience benefit of vehicle-to-home and vehicle-to-grid.

2 - Leveraging Electric Vehicles as a Resiliency Solution for Residential Backup Power During Outages

Shanshan Liu, University of Illinois Urbana-Champaign, Urbana, IL, United States, Alex Vlachokostas, Eleftheria Kontou

Climate change exacerbates power outage events that pose significant risks to local economies and can even endanger citizens' lives. An electric vehicle (EV) equipped with bidirectional energy exchange capability and vehicle-to-home (V2H) technology can serve as a battery resource and provide backup power to meet residential energy needs during an outage, aiming to curb health hazards caused by overheating or cold temperatures. We simulate the electric V2H system in nine US climate regions, over four seasons, and during short-term, long-term, and extremely severe outage events. We propose new resilience metrics to evaluate EV-enabled household energy resilience that measures the duration of time that EVs can serve household energy needs during outages, and mobility resilience, which represents the remaining driving range of EVs after the outage. Our findings indicate that contemporary EV models, even when their state of charge is 50%, can meet residential energy needs during a 12-hour outage in mild seasons (i.e., spring and autumn), except for communities in the Central, West North Central, and East North Central regions due to lower temperatures and higher household energy needs. The resilience of the household energy system during long-term outages is affected by the start time of the outage, heating and cooling power requirements, and daily travel needs. During extremely severe power outages, EVs can protect residents from cold stress over multiple days. Our methods and findings contribute to the coupled management of the integrated residential and EV energy system.We propose strategies for mitigating the adverse impacts of extreme weather events.

3 - Evs&Rsquo; Potential for Smoothing Daily Load Profile Considering Mobility Needs and Solar Output

Nima Rashgi Shishvan, University of Central Florida, Orlando, FL, United States, Zhaomiao Guo

The growth of electric vehicle (EV) adoption and renewable energy sources presents challenges to the power system. Therefore, this study aims to leverage the spatio-temporal charging flexibility of EVs to smooth daily power load and provide ancillary services using vehicle-to-grid (V2G) features. To achieve this objective, we mathematically formulate the aforementioned flexibility based on daily trip chain data and develop a convex optimization model to optimize the smoothness of the regional net load profile, subject to flexibility constraints. We test our approach on real data derived from the Chicago Metropolitan Agency for Planning (CMAP) travel survey and solar generation, as well as net power load data in the area of interest. Our findings indicate that a centralized management system employing the proposed control methodology can significantly smooth the overall power load, considering various levels of charging schemes, EV penetration levels, and infrastructure availability. This study provides an upper bound for future mechanism design to guide the charging behavior of EVs.

WA36

Summit - 428

Modeling and Optimizing Infrastructure for Freight Transportation Operations

Invited Session

TSL: Freight Transportation

Chair: Hyeong Suk Na, University of Missouri, Columbia, MO, 65211, United States

1 - Enhancing Public Transport Evacuation Efficiency: A Round-Trip Bus Evacuation Model

Mina Samiei nasab, University at Buffalo, Buffalo, NY, United States, Rajan Batta

As emergencies such as natural disasters and industrial accidents occur with increasing frequency, the need for effective evacuation planning has become more pressing. Buses, with their large capacity and manageability, are often employed to evacuate individuals without personal vehicles. To enhance the efficiency and practicality of public transport evacuation plans, this study introduces a round-trip bus evacuation model. The model optimizes bus scheduling and routing planning, with constraints including bus capacity, evacuee demand at pick-up points, time windows, and flow balance. The aim is to maximize the total number of evacuees. Additionally, this study emphasizes equity in evacuation planning, ensuring fair attention to all affected populations. To demonstrate the effectiveness of the proposed model and solution, a case study is conducted.

2 - Simulation and Optimization for the Restaurant Food Delivery Problem in Dublin

Luis Cadarso, Rey Juan Carlos University, Fuenlabrada, Spain, Adrian Serrano, Peter Keenan, Javier Faulin

The Restaurant Food Delivery Problem consists of the logistical challenges faced by restaurants in efficiently delivering food orders to customers. This involves optimizing the delivery process to ensure timely and accurate delivery while minimizing costs and maximizing customer satisfaction. Traditionally, restaurants with delivery services rely on their fleet of vehicles, typically motorbikes, which can be costly and troublesome. However, more recently, new delivery systems such as crowdshipping have emerged, offering all restaurants an easy and cost-effective way to handle deliveries. This work aims to assess how a crowd-based bike food delivery system would perform in the city of Dublin (Ireland). To achieve this, an agent-based simulation model is built featuring population, bikers, and restaurants characteristics. Additionally, several optimization problems are integrated into the simulation to guide agents behavior. This also includes a biased-randomization savings heuristic for the Vehicle Routing Problems derived from the traditional own-fleet-based restaurant deliveries. Promising results are obtained for KPI related to economic and customer satisfaction for a set of scenarios based on delivery system, demand rates, and bikers adoption, among other factors.

3 - A Stacked Machine-Learning Method for Predicting Global Supply Chain Restructuring Under Changing Environmental Conditions

Hadi Ghayoomi, George Mason University, Fairfax, VA, United States, Elise Miller-hooks

This talk introduces a stacked ML approach for modeling global supply chains with limited information. Through the identification of the most important features, prediction of how a supply chain might restructure in response to changing logistical opportunities can be made. Such prediction is illustrated on a case study of lithium-ion batteries for possible future scenarios related to changing environmental conditions.

WA37

Summit - 429

Innovations in Urban Transportation

Contributed Session

Chair: Yihan Wang, Tianjin University, Tianjin

1 - A supervised machine learning framework to predict the request fit for dynamic dial-a-ride problems

Simon Mader, KU Eichstätt-Ingolstadt, Ingolstadt, Germany, Pirmin Fontaine, Stefan Voigt

Efficient public transport is key to the sustainable mobility transformation, especially in rural areas where traditional line-based bus services struggle due to the sparse demand. Therefore, innovative on-demand buses, with flexible routes and schedules, present a practical solution for improving rural mobility. Public transport providers aim to maximize the number of served passengers with these on-demand buses but operate within limited resources, leading to challenging decisions on accepting or rejecting passenger requests. Moreover, the dynamic booking process requires an immediate decision without knowing the impact of the request on potential future requests. Therefore, the problem of predicting the request fit for dynamic passenger-maximizing dial-a-ride problems (PM-DARPs) emerges.

To address this problem, we introduce the request fit predictor (RFP) framework. Within this framework, we model the request fit as a binary classification task and learn the fit with supervised machine learning models, having perfect information about historical requests. The RFP combines the request fit learned from static PM-DARPs with dynamic components of the cheapest insertion to guarantee feasibility, integrate request prescience, and facilitate an immediate acceptance/rejection decision.

The RFP is tested on real-life data from a German public transport provider and serves 26.19% more passengers than the current business practice. Compared to the cheapest insertion, the RFP not only serves 7.48% more passengers but also covers 6.68% less distance. In the presentation, we will discuss the trade-offs of the RFP's performance KPIs and give further insights into the framework's fairness and adaptability.

2 - A Nuanced Approach to Project Bus Headway Variations

Xiaotong Ding, MBTA, Boston, MA, United States, John Maleyeff

The presentation describes an approach used by the Massachusetts Bay Transportation Authority (MBTA) to evaluate bus headway performance using a queue modeling approach. Bus headway refers to the interval of time between buses on the same route; its control is crucial for efficient public transit operation. The approach aims to identify patterns that cause headway problems so that strategies can be developed to achieve effective headway performance that minimizes occurrences of long waits for passengers and avoids bus bunching, where multiple buses arrive at a stop simultaneously. The problem was studied by projecting headway variations in consecutive buses' departures based on historical Automatic Vehicle Location (AVL) event based data. The distribution of variations across route lengths, route designs, street types, times of the day and bus schedules will be studied. Various forms of queueing methodologies are used to create projections of headway performance. Based on this work, decisions such as holding the bus at stations, modifying schedules, and controlling traffic signals can be evaluated.

3 - Research on the Emergence and Evolution of Social Stability Risks in Urban Renewal Projects Based on ABM

Yihan Wang, Tianjin University, Tianjin, China, People's Republic of, Hui Sun, Yiyue Zhang, Yuxin Huang, Jidong Chen

Urban development in various countries has entered a stage predominantly focused on upgrading existing assets through urban renewal projects. The key to judging the success of such projects lies in whether residents find their living conditions improved. Significant dissatisfaction among residents can lead to considerable social stability risks. This study delves into the emergence and evolution of social stability risks in urban renewal projects, guided by the theory of risk perception. It posits that public assessment of risks in urban renewal projects is subjectively perceived and characterized by various attributes such as ambiguity. Intuitive fuzzy sets are suggested as a tool to measure this risk perception, with the collective risk perceptions of all individuals forming social stability risks at the project level. The study fully considers the human characteristics of public risk perception, suggesting that significant differences in project characteristics exist. Attributes of the public, complex network features, and modes of perception interaction can influence the transmission and evolution of risk perception. Employing Agent-Based Modeling (ABM) within a simulation research framework, the study simulates various factors and establishes a model for the evolution of social stability risks in urban renewal projects, summarizing patterns observed from multiple projects. This research provides a microscopic research perspective on the emergence and evolution of social stability risks in urban renewal projects, uncovering several evolutionary mechanisms. The findings contribute to the effective management and control of social stability risks in urban renewal projects.

4 - Continuous Approximation for Bipartite Matching with Rideshare Dispatch Application

Bo Jones, United States Naval Academy, Annapolis, MD, United States, Illya Hicks, John Carlsson

We consider matching rideshare drivers and passengers sampled from some continuous distribution in a service region. A pair can be matched if they are within a specified radius of one another. Employing continuous approximation analysis, we provide formulas which predict the number of matches that can be made and characterize this matching cardinality's dependence on the radius. We test our prediction using a simulation on a real road network.

5 - Dual Effects of Ex-Post Regulation in the Gig Economy: A Field Experiment of Gig vs. Traditional Agents in a Call Center

Qiaowen Guo, Washington University in St. Louis, Saint Louis, MO, United States, Fuqiang Zhang, Tianjun Feng

As the gig economy continues to thrive, the shift from ex-ante screening of traditional workforces to ex-post regulation of gig workforces has gained increasing attention. While recent research has highlighted the positive effects of this shift in ride-hailing platforms, its impact on interactive service settings like call centers remains understudied. The interactive nature of call centers requires robust ex-post regulations to control performance and mitigate misconduct among gig workers; however, stringent regulations can have unintended consequences on customers' follow-up actions. Through a 14-day field experiment in a call center partnered with a large e-commerce platform, we study the dual effects of regulations on gig workers' service compared to traditional agents. Our findings reveal that gig workers outperform traditional agents in FCR rate, service duration, and first response time, with increases of 3.49% and decreases of 8.72% and 26.42%, respectively. However, gig workers' risk-averse behaviors resulting from robust regulations, such as transferring cases more frequently, can negatively affect customer repurchase intention. Additionally, we highlight the positive effect of the platform's daily-updated reputation system on enhancing gig workers' service efficiency and confirm the positive impact of consistent schedules on first contact resolution rates for gig workers. We also find that female gig workers exhibit more risk-averse behaviors when faced with a competitive incentive scheme. To help the platform better understand the trade-offs in managing a blended workforce, we propose a two-stage analytical staffing model based on a queueing network. Our case study provides practical insights into optimizing workforce composition for the platform.

WA38

Summit - 430

Aerospace Design, Ethics and Operations with AI/ML

Invited Session

Aviation Applications

Chair: Xin Peng, GE Aerospace, Niskayuna, NY, United States

Co-Chair: Vishwanath Bulusu, Crown Consulting, Inc., Mountview, CA, United States

1 - Machine Learning Application in Aircraft Engine Conceptual Design

Michael Tong, NASA Glenn Research Center, Cleveland, OH, United States

In today's competitive landscape, the effective development and utilization of machine-learning (ML) applications have become imperative across various sectors. This study presents an outline of the procedure involved in creating and implementing ML models for conceptualizing and evaluating aircraft engines. These models leverage supervised deep-learning algorithms to analyze patterns within an open-source repository containing data on both production and research conventional turbofan engines. The main areas of focus encompass crucial engine parameters like thrust-specific fuel consumption (TSFC), engine weight, engine diameter, and turbomachinery stage counts. While the creation of ML models is fundamental for their utilization, ensuring their seamless deployment holds equal significance. To address this aspect, a conversational AI (artificial intelligence) chatbot is constructed, utilizing natural language processing (NLP) techniques, to facilitate the deployment of these ML models. The comprehensive workflow encompasses several key stages: gathering and enhancing engine data, training and cross validating the ML models, testing and evaluating their performance, and finally, deploying, monitoring, and updating the ML models. By following this systematic approach, the aim is to streamline the development and deployment process of ML models tailored for aircraft engine conceptual design.

2 - Engineering the Ethical Use of AI in Aerospace

Paul Ardis, GE Aerospace, Niskayuna, NY, United States

The ethical use of artificial intelligence requires that a system exhibit one key aspect of human behavior: rational, ethically-consistent self restraint. Research turns now to the ability to determine how to establish ethical self-restraint for systems of arbitrary complexity, particularly in the aerospace sector where lives hang in the balance.

3 - Autonomous AAM Operations: Considerations, Challenges, and Opportunities

Pavan Yedavalli, Wisk Aero, Mountain View, CA, United States

In this talk, Dr. Yedavalli will introduce Wisk Aero and the company's position in the Advanced Air Mobility (AAM) landscape, highlighting its plan to deliver the first autonomous eVTOL aircraft to market. He will explain how Wisk models its aviation operations and evaluates market opportunities, introducing their advanced microsimulation platform, called SCOPE. The SCOPE simulation model determines and evaluates a variety of insights, including optimal vertiport locations using unsupervised machine learning, specific air taxi ridership across varying global markets, and detailed autonomous AAM operations to understand highly-trafficked routes, vertiport throughputs, fleet sizes, personnel requirements, and required supporting autonomy infrastructure.

There are also many autonomy considerations at the vertiports, such as precision landing and navigation, that will enable aircraft to safely arrive at the final approach and takeoff (FATO) area on the ground, and command and control (C2) links to enable safe communication across the aircraft, air traffic control, and the multi-vehicle supervisor. Finally, the airspace remains both an enormous opportunity as well as a challenge, as dedicated IFR corridors are expected over time, but will require specific spatial separations in early procedures. The integration of these parameters into the broader SCOPE simulation platform will also be discussed.

Dr. Yedavalli will end the discussion with areas of future exploration and research that can help advance the autonomous AAM industry further.

4 - A UAM Benefit Case Study: the San Francisco Bay Area

Xin Peng, University of California Berkeley, Berkeley, CA, United States

Urbanization and population growth have led to an increase in traffic congestion in many cities worldwide, prompting the exploration of potential solutions such as Urban Air Mobility (UAM). By transporting passengers through the air, UAM can bypass ground-level congestion and reduce the number of vehicles on the roads. This presentation provides a framework for evaluating the value of UAM in alleviating congestion, which includes a UAM vertiport design module for determining vertiport placement and passenger assignment, a GPU-based microstimulator for capturing vehicle movements, and a UAM allocator. The results of a case study conducted in the San Francisco Bay Area demonstrate the potential of multi-modal UAM systems in relieving ground congestion like increasing average travel speeds for some roads, reducing total Vehicle Miles Traveled (VMT) and decreasing travel times for certain ground users. However, the required vertiport infrastructure can be substantial.

WA39

Summit - 431

Advanced Computational and Algorithmic Approaches for Optimization

Contributed Session

Chair: Keyju Lee, Seoul,, Korea, Republic of

1 - GPU Accelerated Branch and Bound through Best First Search

Samiran Kawtikwar, University of Illinois at Urbana Champaign, Urbana, IL, United States, Rakesh Nagi

Branch and Bound has emerged as a fundamental technique for solving optimization problems across various domains due to its ability to systematically explore the solution space to find an optimal solution. Among its many variants, Best First Search stands out for its efficiency in guiding the search towards promising regions. However, the computational demand of exploring such large solution spaces remains a challenge with significant performance and scalability issues. In recent years GPU-accelerated solutions for Branch and Bound have surfaced, but they predominantly rely on less efficient search strategies like Depth First Search or Breadth First Search, due to their ease of parallelization. This work presents the first GPU-accelerated implementation of Branch and Bound employing Best First Search. This couples the computational power of the GPU with the efficiency of Best First Search. Our approach utilizes a GPU resident multi-producer multi-consumer queue design within a single kernel to mitigate communication overheads, yielding a high-performance solution. Testing moderate sized instances of the Resource Constrained Assignment Problem (RCAP) demonstrates a staggering ~20x speedup over CPU. Furthermore, we also show that our parallel implementation mirrors the exploration pattern of Best First Search, approximately matching the number of nodes explored in single threaded implementations.

2 - Results of Fuzzy Re-Sorting in Dynamic Bin Packing

Gene Coffman, Retired - Ford Motor Company, Plymouth, MI, United States, Stephany Coffman-Wolph

Many algorithms for dynamic sequencing leverage bin packing like heuristics that rely on fast and efficient sort algorithms executed multiple times (to reorder jobs by new criteria or because additional items have been added to the queue). The authors will apply a fuzzy sort to a sequencing algorithm and will extend the fuzzy concept to the “fuzzify” the bin backing algorithms.

3 - Analysis of Decoding Logic for the Multiple Yard Crane Scheduling Problem

Keyju Lee, Korea Aerospace University, Goyang-City, Korea, Republic of, Dongjin Jeong, Jungyong Seo, Junjae Chae

The encoding and decoding processes for the solution representation (e.g., chromosome for a genetic algorithm) are crucial for efficiently solving scheduling problems. In this research, a chromosome is encoded with the allocation of tasks to cranes and the task sequence within each crane. Since multiple yard cranes, specifically more than three yard cranes, are in use and certain safety distances must be maintained, cranes can become obstacles to each other during operation. An idling crane sometimes has to give way to another crane to ensure a task can be reached without breaching safety distances. A crane may also have to wait before proceeding to a task location until the other crane completes its task and moves away. A decoding logic capable of creating meticulous evading maneuvers to minimize task waiting time is required. This research focuses on designing and comparing different decoding logics for efficiency.

4 - Stochastic Two-stage Assembly Lot Streaming Problem

Daniel Neira, Virginia Tech, Blacksburg, VA, United States, Subhash Sarin, Manish Bansal

In this paper, we address a stochastic two-stage assembly problem of processing multiple batches of jobs (lots) on the machines in Stage 1 and then assembling the jobs on the machines in Stage 2, under lot streaming. The problem is to determine the sequence in which to process the lots and the equal-sized sublots to use for transferring the jobs from the subassembly machines in Stage 1 to the assembly machines in Stage 2 in order to minimize the makespan. The uncertainty arises in the subassembly times in Stage 1 while the assembly times in Stage 2 are known a priori. The problem is formulated as a two-stage stochastic mixed-integer program with recourse. We propose valid inequalities and also present a polyhedral analysis to study their geometric properties. We also use a multicut L-shaped method in conjunction with a progressive-hedging (PH) based heuristic to solve the problem. Results of a computational investigation are presented that reveal the effectiveness of using the inequalities along with the proposed solution methodologies over the use of Gurobi and SCIP solvers

5 - Precise Bayes Regression

Amin Vahedian Khezerlou, Northern Illinois University, DeKalb, IL, United States

Interpretability in Machine Learning has been the focus of ethical conversations in the research and practitioner communities. Given the black box nature of the Deep Learning methods, today's dominant solution, researchers have proposed interpretable predictive models to compete with the level of accuracy of Deep Learning. The recently proposed Precise Bayes Classifier is an estimation of Bayes Optimal Classifier that provides us with a promising route towards optimal accuracy, while remaining interpretable. Continuing on this path, we propose Precise Bayes Regression. We propose a non-parametric method to directly estimate a probability distribution for the output value, conditioned on the observation of a test point. In cases where evidence is sparse, the model builds a local linear model using the estimations from nearby, non-sparse points. Precise Bayes Regression assumes no functional form for the target function. Moreover, we make no assumptions on the independence of the input variables and no assumptions on their distributions. Through theoretical analysis, we show that with increasing number of samples learned, our error value approaches the irreducible error. Experiments on synthetic data confirm our theoretical findings.

WA40

Summit - 432

Hospital Operations

Contributed Session

Chair: Yuanxin Zhu, Carnegie Mellon University, 927 Maryland Ave, Pittsburgh, 15232

1 - Data-driven decision-making support for health care delivery in the Trauma Center and Emergency Department

Jianing Man, Beijing Institute of Technology, Beijing, China, People's Republic of, Kal Pasupathy

Uncontrolled hemorrhage is one of the leading causes of preventable death among trauma patients, and the early detection of hemorrhage could support rapid decision-making to save lives. Recent use of non-invasive and continuous hemoglobin (SpHb) concentration monitoring tools attracted abundant attention, which could provide real-time hemoglobin measurement compared to delayed laboratory analysis. However, the accuracy is still a concern in healthcare settings. We proposed to use Kernel regression and LOESS models to improve the measurement accuracy by considering its trends and behaviors. Gaussian process and functional regression models are considered to handle the missing values among the continuous measurements and improve the prediction accuracy of lab values. The impact of patient demographics is also discussed.

Besides the physiological indicator and its monitoring, the patient service system also needs close monitoring and analytical support to improve service efficiency and quality. Research opportunities and insights related to the decision-making support of the Emergency Department healthcare service will be shared and discussed, which includes the service system status monitoring and flexible staffing design based on system performance or system status indicators. However, due to the complexity and high uncertainties in the healthcare service system, especially for the ED, it is quite challenging to provide accurate predictions on patient behavior as well as the system status. Challenges and research opportunities will be further discussed as well.

2 - Data-Driven Resource Allocation for Healthcare Facility Optimization Under Uncertainty with SIR Dynamics

Seyedreza Abazari, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Arda Vanli

This paper tackles the formidable challenge of allocating COVID-19 resources in regions with constrained healthcare facilities through a nonlinear mathematical model comprising two key components. Initially, hospital capacities were assessed, and using time series modeling, patient-to-bed ratios were estimated. Subsequently, a Susceptible, Infected, Removed (SIR) model was incorporated to examine disease dynamics, taking into account the dampening effect of allocated resources. The model aimed to minimize unmet demand for beds and transfer costs while managing disease spread within a stochastic framework that captures uncertainties in disease spread and recovery rates. Employing this model to allocate resources across regions, this study leverages data from Florida's 67 counties during the COVID-19 pandemic to enhance our understanding of disease transmission and resource allocation dynamics, particularly focusing on efficient resource distribution in settings limited by uncertainty and scarce resources.

3 - Early prediction of patient admission likelihood in emergency departments

Farzad Zeinali, Clemson University, Clemson, SC, United States, Kevin Taaffe, Chris Gaafary, Ronald Pirrallo, William Jackson, Michael Ramsay, Jessica Hobbs

Efficiently transitioning patients from the emergency department (ED) to inpatient units is a critical challenge in the US, often contributing to ED overcrowding. Early prediction of patients' admission likelihood and the demand for inpatient units enables hospital managers to apply proper strategies before real changes in demand happen. Our developed prediction model incorporates clinical and demographic information, vital signs, triage notes, and external factors like weather conditions and local events. This model serves as a decision support tool to provide hospital managers with near real-time predictions of their unit's demand. The study proves that predicting the likelihood of admitting patients to inpatient units in real-time will help identify the system's changing status and make proactive decisions before any surge in demand happens. Such a responsive decision-making approach can improve healthcare systems' operational efficiency and capacities.

4 - Continuous Remote Patient Monitoring (CRPM) for Post-Discharge Patient Management: A Discrete Event Simulation Approach for Workflow Modeling and Operational Analysis

Yuanxin Zhu, Carnegie Mellon University, Pittsburgh, PA, United States, Rema Padman, Holly Wiberg, Mariana Escallon Barrios, Nirav Shah

Remote patient monitoring (RPM) has gained considerable significance in recent years, especially after the COVID-19 pandemic and Medicare coverage expansion in 2019. However, there has been limited investigation of the workflow integration and operational challenges associated with the deployment of this technology for remote patient management. This study addresses the operational challenge of optimizing resource allocation for patient care in multi-provider workflow planning of continuous remote patient monitoring (CRPM) systems. We use discrete event simulation (DES) modeling within Cascade-HF, a pilot CRPM deployment at a major health system in the US to reduce Heart Failure (HF)-related hospitalizations and readmissions. Physiological data is collected from patients using wearable biosensors over a 30-day period, to determine decompensation risk among HF patients. This is combined with electronic health record (EHR) data and patient-reported symptoms. This limited data is used to create a large synthetic dataset that reasonably mirrors the characteristics of the real dataset. We conduct extensive simulation experiments using the synthetic data to balance resource utilization with differing staffing configurations, operational efficiency metrics and cost with varying patient arrival rates and patient characteristics. The results provide several insights into how a simulation-based approach may be used to support operational decision-making with hospital CRPM systems for optimal staffing and resource allocation strategies based on multiple criteria which depend on the patient arrival rate and the hospital's financial constraints.

WA41

Summit - 433

Stochastic Processing and Matching Networks: Recent Developments

Invited Session

Applied Probability Society

Chair: Sushil Mahavir Varma, Georgia Tech, Atlanta, GA, United States

1 - State-Dependent and State-Independent Assignments in Matching Models with Exchangeable Compatibilities

Rhonda Righter, University of California-Berkeley, Berkeley, CA, United States, Elene Anton, Kristen Gardner, Esa Hyytia, Runhan Xie

We consider a matching system with N classes of drivers (or jobs, etc.) that arrive according to Poisson processes with a common rate and wait to be matched to riders (or service completions, etc.). Riders arrive according to a Poisson process and an arriving rider has a compatibility vector for the N driver classes that is exchangeable across the driver classes. We show that the optimal (response-time minimizing) state-dependent assignment policy is to “assign to the longest compatible queue” (LCQ). This extends classical results with full compatibility. We also show, for the smallest exchangeable matching model with two driver classes and three rider classes (the “M” model), that for state-independent policies, first-come-first-matched (FCFM) performs better than ROM (random order of matching).

2 - The Cost of Impatience in Dynamic Matching

Itai Gurvich, Northwestern University, Kellogg School of Management, Evanston,, IL, United States, Angela Kohlenberg

We study matching queues with abandonment. The simplest of these is the two-sided queue with servers on one side and customers on the other, both arriving dynamically over time and abandoning if not matched by the time their patience elapses. We identify universal, non-asymptotic scaling laws for the matching loss due to abandonment, which we refer to as the “cost-of-impatience.” The scaling laws characterize the way in which this cost depends on the arrival rates and the (possibly different) mean patience of servers and customers. Our characterization reveals four operating regimes identified by an operational measure of patience that brings together mean patience and utilization. The four regimes subsume the regimes that arise in asymptotic (heavy-traffic) approximations. The scaling laws, specialized to each regime, reveal the fundamental structure of the cost-of-impatience and show that its order-of-magnitude is fully determined by (i) a “winner-take-all” competition between customer impatience and utilization, and (ii) the ability to accumulate inventory on the server side. Practically important is that when servers are impatient, the cost-of-impatience is, up to an order-of-magnitude, given by an insightful expression where only the minimum of the two patience rates appears. Considering the trade-off between abandonment and capacity costs, we characterize the scaling of the optimal safety capacity as a function of costs, arrival rates, and patience parameters. We prove that the ability to hold inventory of servers means that the optimal safety capacity grows logarithmically in abandonment cost and, in turn, slower than the square-root growth in the single-sided queue.

3 - Asymptotically Optimal Energy Consumption and Inventory Control in A Make-to-Stock Manufacturing System

Erhun Ozkan, Koç University, Istanbul, Turkey, Barış Tan

We study energy consumption and inventory control in a make-to-stock manufacturing system. We consider a Markovian system with a single server. The server consumes energy, and the associated energy cost depends on the server state: a busy server consumes more power than an idle server does, and an idle server consumes more power than a turned-off server does. However, when a turned-off server is turned on, it must complete a set-up process which is costly and lasts for a while. The objective of the system controller is to minimize the long-run average inventory holding, backorder, and energy consumption costs by deciding when to produce, when to idle or turn-off the server, and when to turn on a turned-off server. We consider the asymptotic regime in which the server is in the conventional heavy-traffic regime. We formulate a Brownian control problem (BCP) which has both impulse and singular controls. Depending on the system parameters, the optimal BCP solution is of either control-band or barrier type. From the optimal BCP solution, we propose a simple heuristic control policy and prove its asymptotic optimality. Finally, we demonstrate the close-to-optimal performance of the heuristic control policy by numerical experiments.

4 - Dynamic Scheduling of a Multiclass Queue in the Halfin-Whitt Regime: a Computational Approach for High-Dimensional Problems

Ebru Kasikaralar, University of Chicago, Booth School of Business, Chicago, IL, United States, Baris Ata

We consider a multi-class queueing model of a telephone call center, in which a system manager dynamically allocates available servers to customer calls. Calls can terminate through either service completion or customer abandonment, and the manager strives to minimize the expected total of holding costs plus abandonment costs over a finite horizon. Focusing on the Halfin-Whitt heavy traffic regime, we derive an approximating diffusion control problem, and building on earlier work by Han et al. (2018), develop a simulation-based computational method for solution of such problems, one that relies heavily on deep neural network technology. Using this computational method, we propose a policy for the original (pre-limit) call center scheduling problem. Finally, the performance of this policy is assessed using test problems based on publicly available call center data. For the test problems considered so far, our policy does as well as the best benchmark we could find. Moreover, our method is computationally feasible at least up to dimension 100, that is, for call centers with 100 or more distinct customer classes.

WA42

Summit - 434

Frontiers in Language Models and Operations

Invited Session

Applied Probability Society

Chair: Tianyi Peng, Columbia Business School, New York, United States

Co-Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

1 - Posterior Sampling via Autoregressive Generation

Kelly Zhang, Columbia University, New York, NY, United States, Tiffany Cai, Hongseok Namkoong, Daniel Russo

Real-world decision-making requires grappling with a perpetual lack of data as environments change; intelligent agents must comprehend uncertainty and actively gather information to resolve it. We propose a new framework for learning active exploration algorithms from massive historical data, which we demonstrate in a cold-start content recommendation bandit setting. First, we pretrain an autoregressive model to predict a sequence of repeated feedback/outcomes (e.g., responses to news articles shown to different users over time). In learning to make accurate predictions, the model implicitly learns an informed prior based on rich features (e.g., article headlines) and how to sharpen beliefs as more outcomes are gathered (e.g., clicks as article is recommended). At decision-time, we autoregressively sample an imagined sequence of outcomes for each action, and choose the action with the largest average imagined reward. Far from a heuristic, our approach is an implementation of Thompson sampling (with a learned prior), a prominent active exploration algorithm. We prove our pretraining loss directly controls online decision-making performance, and we demonstrate our framework on a news recommendation task where incorporating language model text features is crucial for best performance.

2 - More Effective Reward Modeling via Indexed Temporal Difference Learning

Shi Dong, Google DeepMind, Mountainview, CA, United States

Reward modeling in reinforcement learning from human feedback (RLHF) remains an evolving field with open challenges. This talk addresses several limitations with a novel reward model training algorithm that coherently synthesizes information from both pre-trained language models and human-labeled preference datasets. Our approach, which we call indexed temporal difference learning (ITDL), trains an ensemble of reward models to capture complex human preferences and leverages ideas from temporal difference learning, one of the central pillars of reinforcement learning. We use results on simple experiments to demonstrate the algorithm's potential, and leave large-scale empirical studies on real-world language models for future work.

3 - Pre-training and In-Context Learning *IS* Bayesian Inference *a la* De Finetti

Naimeng Ye, Columbia University, New York, NY, United States, Hanming Yang, Andrew Siah, Hongseok Namkoong

In-context learning (ICL) has emerged as a powerful learning paradigm, and a line of work has analyzed this phenomenon through the lens of classical bayesian implicit latent factors/parameters learning. On the other hand, multiple works have studied the importance of quantifying uncertainty in prediction models. In this paper, we aim to address both lines of work by re-interpreting uncertainty-aware-ICL as *explicit* Bayesian inference *a la* De Finetti. From this view, pre-training of ICL optimizing the marginal likelihood of observed sequences *automatically* provides the *correct notion* of uncertainty; compared to fitting priors in conventional empirical Bayes, pre-training fits posterior predictives using transformers. Our observation highlights previously under-explored capabilities of ICL: statistical inference and uncertainty quantification. We theoretically prove that under exchangeability, the usual pre-training loss (perplexity) exactly captures the model's ability to do bayesian inference. We also empirically study the effect of imbuing exchangeability into model design on facilitating uncertainty quantification. On length generalization and coverage, we observe that the imbuing inductive bias improves performance of the model significantly. Finally, we address potential considerations in applying this framework to real-world applications such as language modeling. In particular, choosing the unit of which the data becomes exchangeable is a key consideration in the design of the model.

WA43

Summit - 435

Sajad Khodadadian session

Invited Session

Applied Probability Society

Chair: Sajad Khodadadian, Virginia Tech, Blacksburg, VA, United States

1 - CLT for Two-Timescale Stochastic Approximation with Controlled Markov Noise – Theory and Application to Distributed Stochastic Optimization

Vishwaraj Doshi, NCSU, Charlotte, NC, United States

Two-timescale stochastic approximation (TTSA) is among the most general frameworks for iterative stochastic algorithms. This includes well-known stochastic optimization methods such as SGD variants and those designed for bilevel or minimax problems, as well as reinforcement learning. In this work, we conduct an in-depth asymptotic analysis of TTSA under controlled Markovian noise via central limit theorem (CLT), uncovering the coupled dynamics of TTSA influenced by the underlying Markov chain. This goes beyond previous CLT results of TTSA which only consider Martingale difference noise. We use the TTSA framework to study token algorithms - a family of distributed stochastic optimization algorithms where gradients are sampled by a token traversing a network of agents in random-walk fashion. Typically, these are chosen to be Markov chains that asymptotically sample from a desired distribution. We take a novel approach by replacing the vanilla Markovian token by one which follows a controlled Markov chain - namely the Self-Repellent Radom Walk (SRRW). Defined for any given 'base' Markov chain, the SRRW, parameterized by a positive scalar α, is less likely to transition to states that were highly visited in the past. We prove that the optimization iterate errors of the resulting algorithm converge to zero almost surely, and also prove second-order convergence by applying the TTSA CLT. By deriving the explicit form of the resulting asymptotic covariance matrix corresponding to iterate errors, we show it is always smaller than that of an algorithm driven by the base Markov chain and decreases at rate $O(1/\alpha2)$.

2 - MaxMin-RLHF: Alignment with Diverse Human Preferences

Reinforcement Learning from Human Feedback (RLHF) aligns language models to human preferences by employing a singular reward model derived from preference data. However, the single reward model overlooks the rich diversity of human preferences inherent in data collected from multiple users. In this work, we first derive an impossibility result of alignment with single reward RLHF, thereby highlighting its insufficiency in representing diverse human preferences. Next, we propose to learn a mixture of reward models via an expectation-maximization algorithm and solve a MaxMin alignment objective inspired by the Egalitarian principle in social choice theory to better honor diverse human preferences. We present comprehensive experimental results on small-scale (GPT-2) and large-scale language (with Tulu2-7B)) and show the efficacy of the proposed approach in the presence of diversity among human preferences. We remark that our findings in this work are not only limited to language models but also extend to reinforcement learning in general.

3 - Localized spectral representations for networked reinforcement learning

Zhaolin Ren, Harvard University, Cambridge, MA, United States, Na Li

Networked Markov Decision Processes (MDPs) pose a significant challenge to efficient learning due to the exponential growth of the overall state-action space with the number of agents. Recent works have attempted to address this by leveraging the locally decaying structure of the dynamics to develop localized policies that depend only on a truncated portion of the state-action graph. However, these approaches are limited to finite state-action spaces, leaving the more realistic and challenging scenario of large or continuous state-action spaces in networked MDPs as an open problem. In this work, we leverage recent advances in spectral representation of Q-value functions, which have enabled efficient learning for single-agent MDPs, to derive a networked version of a localized spectral representation for the Q-function of each agent, utilizing the exponential decay structure of network dynamics. Building on these local spectral representations, we design efficient algorithms for continuous state-action networked MDPs. We illustrate the efficacy of our approach with a series of representative experiments.

4 - A Single-Loop Finite-Time Convergent Policy Optimization Algorithm for Mean Field Games

Sujay Bhatt, JPMorgan AI Research, New York City, NY, United States, Sihan Zeng

We study the problem of finding the equilibrium of a mean field game (MFG) -- a policy optimal with respect to the mean field induced by itself. Existing works on this subject take a fixed-point-iteration approach. Due to the lack of strong structure and the non-uniqueness of the equilibrium, they settle for solving a regularized version of the problem, with the regularization weight so large that the equilibrium is the unique fixed point of a contractive mapping. This leads to a regularized solution that can deviate arbitrarily away from the original equilibrium. In this work, we take an innovative approach, based on gradient-based policy optimization rather than fixed-point iteration, to solve the original infinite-horizon average-reward MFG. Besides removing the need to regularize, our algorithm is superior in that it is completely data-driven, single-loop, and single-sample-path. Our first main contribution is to characterize the finite-time convergence of the algorithm to a mean field equilibrium, through a novel multi-time-scale analysis. We support the theoretical results with numerical simulations that illustrate the superior convergence of the algorithm over the prior art.

An MFG reduces to an average-reward Markov decision process (MDP) in the degenerate case where the transition kernel and reward are independent of the mean field. As a byproduct of our analysis for MFGs, we get an actor-critic algorithm for finding the optimal policy in average-reward MDPs with a convergence guarantee matching the state-of-the-art. The prior bound is derived under an unrealistic assumption on the Bellman operator. Our analysis removes the assumption.

WA44

Summit - 436

Appointment Scheduling in Healthcare: From Patients to Physicians

Contributed Session

Chair: Dyutimoy Das, Penn State University, 950 W Aaron Drive, State College, PA, 16803, United States

1 - Fair Appointment Scheduling using Deep Reinforcement Learning

Dyutimoy Das, The Pennsylvania State University, University Park, PA, United States, Soundar Kumara, Paul Griffin

In outpatient settings, an incoming patient often requires multiple sequential appointments to treat their ailments successfully. These appointments may include diagnostic and lab visits or further physician consultation sessions. Though these services are provided within a single healthcare provider network, it is common for the scheduling process to be decentralized with multiple groups given access to the same capacity. This can lead to a lack of communication between different departments and preplanning, which can result in long waiting times for patients and/or an inefficient use of provider resources including equipment and staff. To improve provider efficiency, we propose a centralized appointment scheduling algorithm that keeps track of possible future appointments and considers patient cancellations and no-shows, and ensures that the scheduling process is fair across all patient groups. The sequential appointment scheduling process is modeled as a Markov Decision Process and optimized using an intrinsically motivated reinforcement learning algorithm. The approach is demonstrated for a large provider network in the Midwest.

2 - On the Deterministic Equivalent Formulation of the Appointment Scheduling Problem

Yong-Hong Kuo, The University of Hong Kong, Pokfulam Road, Hong Kong, Mengchuan Zou, Janny LEUNG

In this study, we investigate the appointment scheduling problem (ASP), which involves determining the optimal schedule of appointment intervals for outpatients. We came across a stochastic linear program (SLP) formulated in the literature, which claimed to be equivalent to ASP and was used in subsequent research. However, we demonstrate that this equivalence does not hold in general by presenting a counterexample. To address this issue, we establish a necessary and sufficient condition under which the SLP is equivalent to ASP. Furthermore, we discuss the conditions for equivalence in variations of ASP, such as considering no-show patients and unpunctual patients. To solve the ASP with general cost vectors, we propose a stochastic mixed integer linear program. Additionally, we provide analytical and experimental results for ASP in various settings to further understand the problem and evaluate the effectiveness of our approach.

3 - Improving Stroke Prevention through Enhanced Capacity Planning and Patient Appointment Scheduling

Antoine Sauré, Telfer School of Management, University of Ottawa, Ottawa, ON, Canada, Shahryar Moradi, Jonathan Patrick

Motivated by a problem faced by The Ottawa Hospital (TOH) in Canada, this presentation focuses on the study of patient appointment scheduling practices at a Stroke Prevention Clinic (SPC). A patient who is referred to a SPC is typically scheduled for an initial consult with a neurologist; prior to the initial consult, several tests, depending on the patient's condition, need to be performed. Some SPCs face several challenges in reducing patient wait times for consults and in having all test results available before consultation. To address these challenges, we propose a dynamic multi-priority, multi-resource appointment scheduling model which we approximately solve using Approximate Dynamic Programming (ADP) techniques. The main purpose of this model is to identify policies for allocating available testing and consultation capacity to incoming patients, while reducing patient wait times and increasing the number of test results available before consultation in a cost-efficient manner. The potential benefits from the proposed approach are evaluated by simulating the performance of the resulting policy for a practical example based on data provided by TOH. We also investigate the quality and practical implications of the resulting appointment scheduling policy by comparing its performance to that of other heuristic policies derived from the existing literature on appointment scheduling. We believe the booking guidelines resulting from this research could be used in practice by SPCs to significantly improve stroke prevention and thus decrease the impact of unnecessary delays and incomplete test results on patients' health.

4 - A Comparative Analysis of Queueing Model Accuracy for Healthcare Resource Planning under Epidemic Demand

Alexander Rutherford, Simon Fraser University, Burnaby, BC, Canada, Samantha Zimmerman, Alexa van der Waall, Monica Norena, Peter Dodek

Queueing theory and discrete event simulation models can be used to determine capacity requirements to meet access-to-care targets in acute and critical care resource planning. However, simulation optimization can be computationally intensive in scenarios with heavy demand, and queueing theory approximations may be less accurate when demand is changing rapidly with time—as was the case during the COVID-19 pandemic. To address these challenges, we developed a hybrid optimization approach that combines a simulation-based search procedure with a queuing theory approximation, that was chosen based on a comparative analysis of three different approximation methods. Our analysis compares the pointwise stationary approximation, the modified offered load approximation, and the fixed point approximation under epidemic growth in demand. These methods were evaluated in a case study of mechanical ventilator access in British Columbia, Canada, during the first wave of the COVID-19 pandemic. The case study uses COVID-19 case projections from the British Columbia Centre for Disease Control and critical care data from the British Columbia ICU Database. In the case study results, the fixed point approximation achieved the most accuracy under rapidly growing demand. Our modelling analysis was able to efficiently compute the mechanical ventilator capacity required for 95% of patients to access a ventilator immediately, during the peak of the first wave of the COVID-19 pandemic.

WA45

Summit - 437

Healthcare Operations

Invited Session

Health Applications Society

Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Co-Chair: XInyu Liang, INSEAD, Singapore, Singapore

1 - Brand-Name Versus Generic Drugs: An Empirical Study on Health Outcomes of Generic Atorvastatin Usage

Xinyu Liang, University of Michigan, Ann Arbor, MI, United States

Generic drugs represent over 90% of prescriptions and have saved the US healthcare system over 2.4 trillion dollars over the past decade. The full potential of cost-saving benefits, however, can only be unlocked when the drug's effectiveness is ensured. In this paper, we focus on atorvastatin, the world's best-selling statin designed to alleviate cardiovascular risks, as an example to study its effectiveness by estimating the effect of generic drug usage on health outcomes compared to the brand-name counterpart. We leverage the market entry of generic atorvastatin and address the potential endogeneity using instrumental variables. We find that the usage of generic atorvastatin significantly increases healthcare service utilization with heterogeneous treatment effect cross patient populations, particularly leading to more outpatient visits, emergency department visits, and inpatient admissions. Clinically, despite no significant association with increases in major adverse events and side effects, the generic version leads to higher LDL-C levels and is less effective in mitigating cardiovascular symptoms. Furthermore, we show variations in generic drug effectiveness across manufacturers, particularly between authorized generic manufacturers and regular manufacturers. Our findings corroborate the FDA inspection and recall records that manufacturers identified as producing the least effective drugs also had more quality violations and subsequent product recalls. Our paper highlights the impact of generic drug usage on healthcare outcomes and differential effects based on patient and manufacturer heterogeneity with implications for the FDA, payers, and medical practitioners.

2 - Decoding Patient Trust in AI Gatekeeper Healthcare Systems

Shujing Sun, University of Texas at Dallas, Richardson, TX, United States, Wei Gu, Lauren Lu, Susan Lu

This study investigates the dynamics between humans and algorithms in healthcare, focusing on patient-AI gatekeeper interactions where the AI recommends doctors based on patient-provided symptoms. Through a large field experiment in a healthcare organization, we find a general lack of trust in AI recommendations among patients. Despite this, a positive correlation emerges between patient acceptance and the presence of information on prediction accuracy and peer acceptance. More interestingly, in the presence of both information, patients willingly embrace recommendations with high peer acceptance, even with lower accuracy, suggesting human inclination towards conformity to peer choices more than relying on algorithmic precision.

3 - Understanding the Causes and Costs of Staff Attrition in ICU Nursing

Blair Liu, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Diwas KC, Bradley Staats

This study delves into staff attrition among nurses in Cardiovascular Intensive Care Units (CICUs), with a particular focus on operational factors influencing their decisions to resign. Utilizing a detailed dataset from a comprehensive CICU, we empirically identify the principal drivers of nurse turnover and quantify the subsequent impacts on patient care quality and hospital operations. Our analysis reveals that nurse scheduling, support systems, and burnout prevention influence turnover intentions. The findings advocate for managerial interventions that sustain nursing staff and improve patient outcomes.

WA46

Summit - 438

Towards Resilient Pharmaceutical Supply Chains: Analytical Models for Addressing Drug Quality and Shortage Issues

Invited Session

Health Applications Society

Chair: Sasa Zorc, University of Virginia, Darden School of Business, Charlottesville, VA, 22903, United States

Co-Chair: Jiatao Ding, Nanyang Technological University, Singapore, Singapore

1 - Mitigating Shortages of Generic Drugs: The Role of Reliability Certification

Xiaoyu Wu, Fuqua Business School, Duke University, Durham, NC, United States, Robert Swinney, Can Zhang

Generic drugs play a vital role in keeping costs down in the U.S. healthcare system. However, generic drugs have experienced severe shortages due to supply disruptions, caused by a number of factors, including low profit margins that lead manufacturers to underinvest in supply reliability. Shortages result in numerous negative consequences for patients as well as institutional buyers of generic drugs, such as hospital systems, who may be forced to use more expensive alternative drugs when a shortage occurs. The U.S. Food and Drug Administration has recently sought to address generic drug shortages caused by supply disruptions, and one proposal is to implement a certification system that assesses and discloses the reliability of generic drug manufacturers. In this paper, we investigate the effectiveness of such a certification system. We propose a game theoretic model that captures the effects of generic drug manufacturer reliability certification on manufacturers, buyers, a group purchasing organization, and patients. We find that certification may motivate manufacturers to improve their reliability; however, this may increase patients’ shortage cost once the manufacturers adjust their wholesale prices and the buyers adjust their sourcing strategies in response. Nevertheless, we also identify conditions under which certification can benefit all stakeholders compared to the status quo with no certification. Finally, we find that combining certification with subsidies—where certified reliable manufacturers are subsidized by the government—can further reduce the shortage cost when the subsidy level is sufficiently high, but it may backfire if the subsidy level is too low.

2 - Reimbursement Policy and Drug Shortages

Xuejun Zhao, UNC Charlotte, Charlotte, NC, United States, Justin Jia, Hui Zhao

Drug shortages have posed a significant challenge for the U.S. pharmaceutical industry and the government in recent years, causing severe consequences and drawing widespread attention. Substantial efforts have been devoted to identifying the causes, with one debate linking the shortages to Medicare's adoption of the Average Sales Price (ASP) policy for hospital drug reimbursement. Few studies have captured the main tradeoffs and nuances in drug wholesale pricing decisions and subsequent shortages under the reimbursement policy, leaving the exact link between the policy and drug shortages unclear. Recognizing that the reimbursement policy influences shortages through affecting supply chain parties' decisions and given the scarcity of data on GPOs, who are pivotal in these decisions, we analyze a drug supply model that captures the essential elements and tradeoffs in drug wholesale price decisions under the reimbursement policy. We find that under the ASP policy, the interplay of two opposing effects, the free-ride effect and the coordination effect, guides wholesale pricing decisions and affects shortage outcomes. We capture key factors influencing these effects and show that the ASP policy actually possesses resistance to shortages of the drugs that have experienced shortages. We further conduct numerical analysis incorporating real pharmaceutical data (with the aid of machine learning) to gain additional insights. The numerical analysis confirms the ASP policy's resistance to drug shortages and shows that overemphasis on lowering wholesale prices through GPOs can exacerbate drug shortages. Finally, we provide a thorough discussion of policy implications based on our findings.

3 - Impacts of a Non-Profit Intermediary on Reliability of Drug Supply

Junghee Lee, University of Notre Dame, Notre Dame, IN, United States, Jennifer Ryan, Hyoduk Shin, Daewon Sun

Generic drug shortages have become increasingly common and costly in the U.S. Recently, non-profit organizations dedicated to enhancing the reliability of the drug supply have begun to enter the pharmaceutical supply chain, alongside traditional for-profit intermediaries, such as group purchasing organizations. In this paper, we study the impact of a non-profit intermediary on the reliability of supply in the context of a single drug that is vulnerable to shortages. We develop an economic model tailored to long-off-patent generic drugs, produced by a limited number of suppliers. We find that the non-profit intermediary's entry will improve supply reliability for its members but will not necessarily improve the overall supply chain reliability. However, in a long-term setting, in which both suppliers can adjust their wholesale prices, for some cases, more providers will see improved reliability of supply. Surprisingly, we also discover that an increase in the reimbursement rate for the vulnerable drug may not lead to more reliable supply. Our results provide critical insights to government officials and policy makers given that increased reimbursement rates for generic drugs are often suggested as one possible solution to the drug shortage crisis. Our results also provide useful guidance for non-profit organizations who actively seek to mitigate drug shortages. We find that such efforts may have limited impact on the reliability of the drug supply if not properly implemented and that decisions such as the degree of vertical integration can have a significant effect.

4 - Optimal Conditional Drug Approval

Chenxi Xu, Duke University, Durham, NC, United States, Giuseppe Lopomo, David Ridley, Peng Sun

New prescription drugs require regulatory approval before drug makers can sell them. In some countries, regulators may conditionally approve a drug, which allows sales to begin before the developer has proven the drug's efficacy. After further testing, the regulator may either grant final approval or reject the drug. We show that conditional approval not only speeds access to drugs but also encourages the development of drugs that would not have been pursued otherwise. Using mechanism design principles, we show that regulators should conditionally approve a drug even if it is ex-ante less likely to prove efficacious, under certain conditions. The regulator should approve the drug to encourage investment, especially when only the firm knows the testing cost. However, drugs that are less likely to prove efficacious should only be conditionally approved for a portion of the patient population if that is enough to motivate testing. Additionally, the impact of conditional approval is greater when the firm's revenue increases over time. Finally, a regulator should sometimes commit that in the future it will grant final approval for a drug that narrowly misses the efficacy threshold in return for the firm testing an otherwise unprofitable drug.

WA47

Summit - 439

Healthcare Operations

Invited Session

Health Applications Society

Chair: Sindhoora Prakash, University of Massachusetts Amherst, Amherst, MA

Co-Chair: Hari Balasubramanian, University of Massachusetts, Amherst, Amherst, MA, United States

1 - Scheduling Interventions for Individuals in Complex Care Programs

Sindhoora Prakash, University of Massachusetts Amherst, Amherst, MA, United States, Hari Balasubramanian

Patients with multiple chronic conditions, as well as unmet social needs, have significantly higher healthcare costs and utilization. Holistic, person-centered care interventions, often led by a multidisciplinary team consisting of nurses, community health workers, and social workers, have emerged as a strategy to engage with and help improve the health and well-being of such patients. The objective of these interventions is to reduce the occurrence of unnecessary hospitalizations and overall healthcare costs. In this talk, we highlight the potential benefits of using simulation and mixed integer programming to estimate and optimize the allocation of health and social care services to patients, such that their waiting times are limited while minimizing the variance of the number of staff hours required per week.

2 - Outpatient Appointment Scheduling in the Multi-Physician Setting

Haolin Feng, Sun Yat-sen University, Guangzhou, China, People's Republic of, Zitian Li

We study outpatient appointment scheduling in a multi-physician setting with stochastic service times. The decisions are the appointment times for the set of patients given. The objective is to optimize the weighted expected value of patient waits, physician idling, and overtime. We derive expressions for the performance measures and provide theoretical results of the optimization problem. Numerical studies are included to shed light on the managerial implications.

3 - Novel absence forecasting for nursing resource team sizing using nurse-specific absence probabilities

Taylor Martin, University of Toronto, Toronto, ON, Canada, Michael Santorelli, Michael Carter

This study introduces a novel approach to predict nurse absenteeism in an acute care setting. Results are used to plan optimal replacement staffing, such as a Nursing Resource Team. The method, unique in its use of individual staff characteristics, offers improved accuracy compared to traditional average-based approaches. Using historical attendance data to train a machine learning model, individual likelihoods of absence are predicted for each upcoming shift and an expected value across all staff is computed as the forecast for each shift. The novelty of this approach is that for each shift the predicted number of absent nurses is influenced by the individuals scheduled to work. Forecast absence data are used in an optimization model to determine levels of pre-booked and ad-hoc staffing under operationally relevant constraints. Testing over 244 weeks of real data from an acute hospital indicated that the optimized machine learning model produced total staffing costs that were 5% lower, with 15% less regret, than baseline. Testing revealed superior performance when i) nurse-specific identifiers were included (‘nurse-aware’ models) and ii) more training data were used. This research demonstrates the potential for nurse-aware approaches to forecasting and optimization in an applied setting. Optimization results demonstrated that an ML approach yields lower numbers of pre-scheduled staff and higher values of ad-hoc straight time shifts. What does mean for the workforce? In the current global nursing shortage lower staffing requirements are attractive, and optimization leads to fewer nurses being better utilized.

4 - Maintaining Fairness in Chemotherapy Scheduling Under Uncertainty

Batuhan Celik, Rennselaer Polytechnic Institute, Troy, NY, United States, Serhat Gul, Ozlem Karsu

Chemotherapy scheduling is hard to manage under uncertainty in infusion durations, and focusing on expected performance measure values may lead to unfavorable outcomes for some patients. In this talk, we aim to design daily patient appointment schedules considering a fair environment regarding patient waiting times. We propose using a metric that encourages fairness and efficiency in waiting time allocations. To optimize this metric, we formulate a two-stage stochastic mixed-integer nonlinear programming model. We employ a binary search algorithm to identify the optimal schedule, and then propose a modified binary search algorithm (MBSA) to enhance computational capability. Moreover, to address stochastic feasibility problems at each MBSA iteration, we introduce a novel reduce-and-augment algorithm that utilizes scenario set reduction and augmentation methods. We use real data from a major oncology hospital to show the efficacy of MBSA. We compare the schedules identified by MBSA with both the baseline schedules from the oncology hospital and those generated by commonly employed scheduling heuristics. Finally, we highlight the significance of considering uncertainty in infusion durations to maintain fairness while creating appointment schedules.

WA48

Summit - 440

EVs -- Adoption, Operations and Interface with the Utility Grid

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Siddharth Singh, University College London, London, United Kingdom

1 - Utilities' Managed Home-Charging Programs for Electric Vehicles

Ali Fattahi, Johns Hopkins University, Baltimore, MD, United States

Experts estimate 20 million electric vehicles (EVs) will be on U.S. roads by 2030, and the majority (around 80%) of EV drivers will use home charging. Many utilities are designing managed home charging (MH) programs to centrally manage EV drivers' home charging to reduce cost, avoid new and aggravated peaks and blackouts, and ensure grid stability. An MH program is either an active program (AMH), in which the utility continuously controls the EV charging while the vehicle is plugged in, or a passive program (PMH), in which the participating EV drivers decide when to charge, based on pre-announced low-rate episodes. We present a program-design model (PD) and a load-management model (LM) for jointly designing and executing MH programs. PD designs a menu of MH programs, tailored for each driver type, and LM dynamically manages the load supply to each individual participant. LM consists of a large number of non-homogeneous participants, and it is a large-scale mixed-integer nonlinear stochastic problem. We present an effective approximation method, conduct thorough theoretical and numerical analyses of our approximation, and provide worst-case bounds for its error components. Our methodology provides detailed insights on the amount and timing of the improvements achievable in cost and demand variability by offering AMH, PMH, and both, and by customizing PMH. It also offers detailed insights on the significance of the trade-off between cost and demand variability. We find promoting a culture of charging EVs every night may significantly increase utilities' total cost if PMH has a high participation level.

2 - Optimal Megawatt Charging Station Investment and Design: Mitigating Peak Demand

Ridvan Aksu, The University of Alabama, Tuscaloosa, AL, United States, Mesut Yavuz

Medium and Heavy Duty (MDHD) trucks contributes the 6% of the greenhouse gas emissions of the US while accounting for the 29% of the transportation fuel use. In order to adapt Electric MDHD, an established fast charging network is required, given their high energy demand, current limitations of battery and charging technologies, and the time sensitive nature of land freight transportation. Existing plans for Megawatt Charging Systems (MCS) and zero-emission freight charging corridors to expedite this transformation requires a heavy load on grid networks, given the power demand of each megawatt charging port being the size of a small to medium size manufacturing plant. In this work, we investigate utilizing on-site renewable energy generation and energy storage systems to mitigate the peak load on the grid and providing a quick turnaround time for MDHD vehicles while having a profitable business plan, given the high investment cost of charging, storage, and energy generation systems. We also analyze charging policies that balances the tradeoff of the grid load and customer satisfaction with respect to customer arrival density and energy pricing.

3 - Watt's Next: Navigating the Surge in Electric Vehicle Adoption and its Implications for Grid Operations

Siddharth Singh, University College London, London, United Kingdom, Karthik Murali, Owen Wu, Mesut Yavuz

As policy decisions pave the way for more Electric Vehicles (EVs) on the road, there is expected to be a tremendous strain on the utility grid. Managing this strain requires careful planning for (and possibly control of) EV charging operations by the utility firm. This may include the utility firm managing intra-day demand variations using grid-scale batteries, and taking over charging operations of some fraction of EVs (via a managed charging program), including the possible use of Vehicle-to-Grid (V2G) technologies. A managed charging program must be carefully designed, accounting for EV users' driving patterns and range anxieties. Using a stylized analytical model that considers two periods per day (peak and off-peak), we derive optimal grid operations in the presence of a fraction of EV owners who sign up to a managed charging program. We find the optimal operating policy and show that both V2G and grid-scale battery usage are only useful when the fraction of managed EVs is low. We demonstrate the validity of all our findings numerically using a higher fidelity hour-to-hour model calibrated to real data. Our results suggest that a long-term outlook on navigating EV adoption should focus on managing more EV charging, rather than installing stationary battery capacity or promoting V2G. Our results also carry useful implications for policy-makers on the optimal target level of EV adoption under various objectives that weigh environmental factors differently.

4 - Economic Potential and Benefits of PV-Integrated Vehicle-to-Grid (V2G) Technology in the Rent-a-Car Industry: A Case Study of Jeju Island

Yoonmo Koo, Seoul National University, Seoul, Korea, Republic of, Wonjong Lee, Jonghyeok Han

Vehicle-to-grid (V2G) technology, enabling bidirectional power transactions between the power grid and electric vehicles (EVs), is recognized as a promising solution for the deep integration of EVs with renewable energy (RE). While substantial research has been conducted on V2G applications for private EVs and car-sharing businesses, the exploration of V2G within the rent-a-car industry remains limited. This study investigates the economic potential and advantages of a PV-integrated EV rent-a-car business in tourist destinations, with a particular focus on Jeju Island, a leading tourist spot in South Korea.

Utilizing 105,473 rental and return data points from 2023, this research employs bilevel programming to simulate the decision-making and interaction between the rent-a-car business and the system operator. The upper level models the V2G/G2V bidding strategies of the rent-a-car business, while the lower level addresses the market-clearing decisions of the system operator. By applying the Karush-Kuhn-Tucker conditions and the strong duality theorem, the model is transformed into a single-level mixed integer linear programming problem.

The findings indicate that an optimal configuration—15 MW of solar PV, 10 MW of unidirectional chargers, and 30 MW of bidirectional chargers—can generate an annual revenue of USD 3,205,580, accounting for investment and battery degradation costs. Additionally, the power system's operation costs were reduced by up to 6.8% through the alleviation of RE curtailments by 5.5%. These results highlight a novel revenue stream and co-benefits for V2G in the EV rent-a-car business in tourist areas, suggesting a feasible decarbonization strategy to mitigate carbon emissions in such locales.

WA49

Summit - 441

Collaboration in Supply Chains with Channel Partners and Competitors

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Abhishek Roy, Temple University Fox School of Business, Philadelphia, PA, 19122, United States

1 - The First-Mile Problem in the Reverse Supply Chain of Used Electronics

Yunke Mai, University of Kentucky, Lexington, KY, United States, Haoying Sun

In this paper, we study how to incentivize consumers to send their end-of-life electronics for reuse and recycling under current reverse supply chain structures. We show that utilizing existing retail channels could be an effective approach and establish the optimal subsidy policies that increase the recycling rate.

2 - A Random Model of Supply Chain Networks

Philippe Blaettchen, Bayes Business School (formerly Cass), London, United Kingdom, Andre Calmon, Georgina Hall

Many optimization problems in supply chain management are formulated over graphs representing networks of interlinked supply chains. Solution approaches to these problems are often studied and tested on small, stylized supply chain networks, or on some of the few publicly available supply chain network dataset. This paucity of data means that fundamental questions about the relationship between the network structure and the properties of the problem solution such as quality and complexity are not addressed systematically.

Our paper fills this gap by introducing a random generative model of supply chain networks, in line with the computer science and social sciences literature, where random graph models are a popular tool for studying properties of “typical” networks. We show that our model, based on simple micro-foundations, generates network structures similar to those observed in practice. We propose that it supports (i) analyzing how network structure affects computational complexity, (ii) identifying new managerial insights, and (iii) benchmarking heuristics. We illustrate these benefits with a case study on safety stock optimization.

3 - Progressive referral fee on online retail platforms

Aadhaar Chaturvedi, University of Auckland, Auckland, New Zealand, Aditya Jain

To boost product variety, online retailers have adopted agency selling model wherein the supplier determines prices and retailer charges a fraction of price as referral fee. This paper tackles the important question of how should a retail platform design progressive referral fees facing heterogenous suppliers differing in demand and cost.

4 - Impact of cocreation under supply chain encroachment

Abhishek Roy, Temple University Fox School of Business, Philadelphia, PA, United States, PIYAL SARKAR

The increasing complexity of supply chains has led firms to redesign supply chain strategies based on the collaborative value created through co-creation with other supply chain members. Co-creation is an economic strategy that binds supply chain members to jointly improve the output. Our research investigates how co-creation can add value under supplier encroachment environment. Encroachment refers to the scenario when a manufacturer establishes a direct sales channel (especially via an online platform). Supplier encroachment mitigates double marginalization and secures Pareto-improvements for both parties. Companies such as Apple, Nike, Adidas practice cocreation as well as encroachment strategies. Previous literature has investigated the impact of cocreation under different settings, but the impact of cocreation under supply chain encroachment environment has received little attention. This study addresses this gap and tries to draw managerial insights by integrating the co-creation and encroachment strategies. The study aims to provide the following insights: a) co-creation strategies given quality is the main differentiator of the product under supply chain encroachment environment and b) encroachment strategies given co-creation exists between the manufacturer and retailer.

WA50

Summit - 442

Care Access and Delivery

Invited Session

MSOM: Healthcare

Chair: Anita Carson, Boston University, Boston, MA, United States

Co-Chair: Ruozhu (Savannah) Wang, Boston university, Boston, MA, 02215, United States

1 - Work Design and Scheduling for Dialysis Clinics

Vincent Slaugh, Cornell University, Ithaca, NY, United States, Andre Cire

Based on work with a large network of dialysis clinics, we show potential improvements for labor costs and work design by switching from a "push" schedule to an "interlaced" schedule. We analyze closed-form expressions of day length and consider tandem work models. We also share insights from implementation.

2 - When Is Standardization Most Beneficial for Improving Healthcare Service Quality? The Moderating Role of Operational Failures

Qi Wang, Boston University, Boston, MA, United States

Research finds a mixed result on clinical outcomes from standardizing medical procedures. Under some conditions, standardization improves performance, while under other conditions, it hurts. We investigate a condition we term “operational failures” – influences the relationship between standardized procedures and clinical outcomes. We conduct two separate studies to test whether operational failures moderate the relationship between standardized procedures and clinical outcomes. Study 1 uses survey data from nurses in medical and surgical units across 56 U.S. hospitals to gather measures of operational failures and standardization and matches them with data on pressure injuries. Study 2 triangulates Study 1 by using objective clinical data from over 17,000 patient visits at a major U.S. medical center. Through both studies, we find strong support that standardization enhances clinical outcomes in units that suffer from frequent operational failures. We suspect that standardization is most beneficial when it provides structure in a chaotic work environment. Conversely, in units that experience few operational failures, standardization provides little benefit for clinical outcomes. This surprising result may be because of the downsides of standardization: it limits clinicians’ flexibility to respond to patients’ unique needs. In a reliable work environment, there is less need to impose the rigidity that comes from highly standardized procedures. We provide insight to healthcare managers by explaining that standardization’s positive impact peaks in environments with frequent operational failures but offers little benefit when operational failures are rare.

3 - Seeing is Believing - Surgery Video Sharing Improves Trust , Adherence and Health Outcomes

yueran Hou, Southwestern University of Finance and Economics, Chengdu, China, People's Republic of, Lin Zhao, Meng Li

In the realm of healthcare, the significance of trust cannot be understated. It serves to mitigate feelings of vulnerability in patients and enhances the efficacy perceived by clinicians. The patient-physician relationship is often presupposed to be underpinned by trust due to its fundamental nature. However, with the evolution of healthcare systems and societal shifts, trust is increasingly understood to be at risk and in need of attention, medical disputes arising from mistrust are becoming more common. In this paper, We have proposed a method to bolster the trust between physicians and patients - sharing surgical videos with patients. We have conducted research on the impact of this approach on the psychological, physiological, and behavioral aspects of patients.

4 - The Effect of Must-Access Prescription Drug Monitoring Programs on Opioid Prescriber Rates

Behrooz Pourghannad, Lundquist College of Business, University of Oregon, Eugene, OR, United States, Beyza Celik, Guihua Wang

We examine the effect of must-access Prescription Drug Monitoring Programs (PDMPs) on opioid prescriber rates. We first use the difference-in-differences approach to analyze the average effect of PDMPs. We then use an empirical machine learning approach to analyze the treatment effect heterogeneity across healthcare providers. Our analysis sheds light on the effects of must-access PDMPs, informing policymakers and stakeholders about their implications in addressing the opioid crisis and optimizing patient care.

WA51

Summit - 443

Data-Driven Methods to Human - Artificial Intelligence Interactions

Invited Session

MSOM: Service Operations

Chair: Agni Orfanoudaki, Oxford University, Oxford, OX26SQ, United Kingdom

1 - Generative AI at Work: Hallucination, Moral Hazard, and Contracting

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States, Terry Taylor

We examine how generative AI can affect workplace dynamics, with a particular focus on how the phenomenon of hallucination can complicate the incentive design problem. We show that moral hazard may be either exacerbated or mitigated by the hallucination of generative AI outputs. Our findings point to the need to rethink incentive design to facilitate effective human-AI interaction in the workplace.

2 - Algorithm-in-the-Loop: Towards Better Human Decision Making

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States, Agni Orfanoudaki

As data availability continues to expand, the deployment of data-driven algorithms to assist human decision-making is becoming increasingly prevalent across a wide range of industries. However, the effectiveness of these algorithms can vary significantly depending on how human decision-makers are influenced by the algorithmic advice. This paper establishes a comprehensive framework for measuring and optimizing the positive impact of algorithms on human decision-making.

We introduce the concept of an influence function, which is central to understanding human performance when algorithmic advice is available. Different behavioral assumptions can be naturally encoded within the influence function, allowing us to capture various modalities of algorithmic benefits, ranging from harmful and contradictory to complementary and beneficial.

To demonstrate the practical applicability of our framework, we conducted a randomized study in a large academic hospital. This study involved doctors and focused on the impact of algorithmic advice on the prescription of antibiotic prophylaxis. Our results showcase the framework's ability to accurately measure the heterogeneity of influence functions across different doctors and personalize algorithmic outputs to optimize human performance.

3 - Towards Stable Machine Learning Model Retraining via Slowly Varying Sequences

Vassilis Digalakis, HEC Paris, Jouy-en-Josas, France

Stability is important for both ensuring consistent performance and enhancing interpretability: significant changes in a model’s structure or the resulting analytical insights can lead to skepticism and hesitation for adoption. In this talk, I will present my research in the area of stable and “slowly varying” ML through the lens of healthcare applications and collaborations with major US hospitals. First, I will introduce the framework of slowly varying regression under sparsity, allowing sparse regression models to exhibit controlled, slow and sparse variations. Next, I will propose a decision tree stabilization methodology relying on a new distance metric for decision trees, which is used to determine a tree's level of stability. Lastly, I will discuss a general, model-agnostic framework aimed at stabilizing ML model structures (including black-box models such as boosted trees) upon retraining with new data. We have tested the proposed methodologies on a variety of real-world case studies and have shown, empirically and theoretically, that, with a small, controlled decrease in predictive power, we often gain a significant improvement in the models' stability.

WA52

Summit - 444

Empirical Research on Marketplaces

Invited Session

MSOM: Service Operations

Chair: Rakesh Allu, Cornell University, Ithaca, NY, United States

Co-Chair: Vishal Gaur, Cornell University, Ithaca, NY, 14850, United States

1 - Causal Product Networks: Discovery and Applications for Basket Shopping

Ziwei Zhu, Cornell University, Ithaca, NY, United States, Vishal Gaur, Nur Kaynar

Modeling the purchase decisions of basket-shopping consumers is a hard problem. In this paper, we employ a novel causal structure learning technique to uncover the causal relationships among product purchases and study two main research questions: (1) How do causal product networks perform in describing the relationships among product purchases when compared to various other network structures that represent hypothetical consumer behavior? (2) Given the empirical evidence for network structures, how can we construct the optimal assortment strategy based on the causal relationships among product purchases, and to what extent does it outperform a benchmark within-category assortment strategy based on Multinomial Logit (MNL) models? Our empirical evidence shows that across all five themes, the model based on the causal structure among products consistently provides the best fit. Besides, we compare the performance of causal product networks and MNL models in both brick-and-mortar and online channels, and find that the assortment strategies derived from the former significantly outperform those from the latter, achieving objective values that are two to three times higher. Thus, our study yields insights into consumer behavior as well as provides a parsimonious model for estimating basket-shopping behavior from historical data and thereby optimizing assortments across product categories in an integrated model.

2 - The Role of Algorithmic Housing Market Intermediary in Attenuating Racial Price Differentials - Evidence from iBuyers

Yuan Cheng, Cornell University, Jersey City, NJ, United States

Algorithms powered digital technologies are playing an increasingly important role in automating home buying process, offering efficiency and convenience for home buyers. However, it also raises the fairness concerns, as algorithms not designed to discriminate can also discriminate due to bias in the data. In this paper, we examine whether a new type of housing market intermediary -- iBuyers, which rely heavily on algorithms to compute housing valuations -- exacerbate or attenuate racial price differentials in the US housing market. Using millions of housing transactions and mortgage data and a repeat-sales framework that ensures comparable housing, we find strong evidence that iBuyers attenuate existing racial price differentials in markets with the most iBuyer presence. Our findings are robust to the potential unobserved changes in housing characteristics, which are achieved by using coarsened exact matching. For comparison, we further investigate the traditional housing market intermediary, known as flippers, but find no evidence of similar function as iBuyers in addressing racial disparities. Finally, our heterogeneity analyses shed light on the mechanism of iBuyers' role in attenuating racial price differentials.

3 - Do Online Marketplaces Democratize Market Access? Evidence from a B2B Platform.

Vishal Gaur, Cornell University, Ithaca, NY, United States

TBD

4 - Shift Commitment in the Gig Economy: A Study with a Food Delivery Platform

Yao Cui, Cornell University , Ithaca, NY, United States

Online delivery platforms often rely on gig workers to serve as delivery drivers. While there are certain purported benefits to utilizing gig workers for this role, it often leads to a challenge in maintaining an adequate supply of drivers to meet the fluctuating customer demand throughout the day. Hence, to address this issue, platforms have been designing new incentives schemes to encourage greater commitment in the availability of the gig workers at the crucial hours. In this study, we are interested in how platforms can resolve the demand and supply imbalances through incentivizing the gig workers to commit to certain shifts. Using a comprehensive dataset from a major food delivery platform in Asia, we empirically study the impact of shift commitment on gig workers' behavior, and provide guidance for platforms seeking to enhance the reliability of gig worker availability.

WA53

Summit - 445

Food Supply Chain Management

Invited Session

MSOM: Sustainable Operations

Chair: Stanley Lim, Michigan State University, East Lansing, MI, 48824, United States

Co-Chair: Goutham Takasi, The University of Alabama, Tuscaloosa, AL, 35406, United States

Co-Chair: Goutham Takasi, The University of Alabama, Tuscaloosa, AL, 35406, United States

1 - Reducing Food Waste: The Impact of Online Grocery Shopping

Jimin Park, Massachusetts Institute of Technology, Cambridge, MA, United States, Yanchong Zheng

The use of online grocery shopping has been rapidly increasing in recent years. With food waste reduction becoming a global priority, it is critical to understand the impact of online grocery shopping on food waste. This paper examines how the advent of online grocery shopping influences household and retail-level decisions and analyzes its role in reducing food waste along the supply chain system. We develop a game-theoretic model to characterize a two-level grocery supply chain, comprised of the household and the retail-level, under the base scenario of no online shopping. We examine how product and consumer characteristics impact the decisions and the waste outcomes. We then analyze and contrast the model with the adoption of online shopping with the base scenario to uncover its differential impacts on the equilibrium. Finally, we use actual transaction data from the US to empirically validate our modeling insights and quantify the impact of online grocery shopping on food waste. Our findings reveal that online grocery shopping has the potential to decrease food waste at both the household and retail-levels, but its effectiveness depends on being adopted by the "right" households and product lines. Specifically, online shopping successfully reduces food waste at the household-level when utilized by households who are distant from the retailer on products that are less perishable and are consumed at a slower rate. Analytical and empirical results further demonstrate that retail-level food waste always decreases under the same conditions due to the increased power in managing inventory.

2 - Variability Propagation in Perishable Product Supply Chains

Xiaoyue Yan, University of Oregon, Eugene, OR, United States, Elena Belavina

Demand variability and its propagation in the supply chain have played a key role in recent shortages, inflation and turmoil. Managing demand variability is essential to minimizing costs and delivering reliable supply. In perishable-product supply chains it is also key to reducing food waste and carbon emissions. This study provides the analysis of variability propagation in perishable-product supply chains. We build and calibrate a two-echelon perishable-inventory model, showing that the nature of variability propagation in perishable-product supply chains is strikingly different from that in well-studied durable-product supply chains. In particular, we find that (i) product perishability is a novel, hitherto unknown driver of the much-examined bullwhip effect (upstream variability amplification), (ii) surprisingly, perishability can also lead to upstream variability attenuation, an *anti*-bullwhip effect. Our data-driven model calibration reveals a great variation in the degree of variability amplification across different products resulting from more/less favorable combinations of the product and market characteristics. Products with more extreme (high or low) purchasing price, replenishment cost, mean of product expiration time, and standard deviation of buyer demand exhibit higher variability amplification. High mean of buyer demand and standard deviation of product expiration time also yield higher amplification. Finally, we show that the buyer's order quantity modulates the extent of the upstream variability amplification, and as a result, the supply chain partners could attempt to identify contracts that coordinate on buyer's order quantities to limit variability amplification. This could lead to overall less food waste (3-6%) and higher profits (2-10%) for the supply chain.

3 - Farm Equipment Sharing in Emerging Economies

Priyank Arora, University of South Carolina, Columbia, SC, United States, Olufunke Adebola, Can Zhang

In emerging economies, there is a growing number of farm equipment sharing platforms that connect smallholder farmers with tractor owners who are willing to fulfill farmers' requests for mechanization services. Due to the small farm sizes and the low digital literacy in rural areas of emerging economies, these platforms often rely on the so-called "booking agents" to collect demand from individual farmers and submit the aggregated demand on the platform (rather than having individual farmers submit their service requests). This paper studies how the presence of such booking agents affects the platform's optimal pricing and wage decisions and the equilibrium outcomes. Our analysis reveals several insights with managerial implications. First, in contrast to conventional sharing settings, we show that in the presence of booking agents, an increase in the number of service providers (i.e., tractors) may lead to a lower optimal platform commission. Second, while supply-side interventions, such as increasing the number of tractors on the platform, generally lead to a decrease in price and an increase in total farmer surplus, we find that reducing booking agents' demand aggregation cost can lead to an increase in price and a decrease in total farmer surplus. Finally, although the platform needs to pay booking agents when it relies on these agents to collect demand, we show that the presence of these agents leads to a "win-win-win" outcome for farmers, tractor owners, and the platform when booking agents' demand aggregation cost is moderate or the number of tractors exceeds a certain threshold.

4 - Enhancing Agri-Food Supply Chain Resilience: Implications for Wholesale Prices and Inventory Management

Goutham Takasi, University of Alabama, Tuscaloosa, AL, United States, Stanley Lim, Monireh Mahmoudi

This study examines the operational challenges local food distributors encounter, particularly when sourcing from small and medium-sized farms (SMFs), which significantly impact the supply chain's resilience. The inconsistency in supply from SMFs, attributed to factors such as limited capacity and inadequate infrastructure, poses a risk of supply chain disruptions and potential retail shortages. To address these issues, we collaborate with Tamarack Holdings (TH), a Michigan-based agri-food distributor, with the aim of enhancing supply chain resilience through improved demand estimation and strategic inventory management. Our research is twofold: firstly, we develop a choice-based analytical framework to accurately estimate product demand, considering the multifaceted nature of customer orders in terms of the variety and quantity of stock-keeping units (SKUs). This model facilitates the identification of critical SKUs, enabling TH to prioritize resources and design effective farmer incentive schemes, potentially through adjustments in wholesale pricing. Utilizing a unique dataset comprising 3,292 SKUs over a 12-month period in 2022, which includes daily inventory data, replenishment orders, and customer transactions, we employ the multiple discrete-continuous extreme value (MDCEV) choice model to capture customer purchasing behaviors accurately. Our approach aims to provide insights into product consumption and substitution patterns, thereby contributing to the literature on supply chain management and inventory control in agri-food distribution.

WA54

Summit - 446

Advancing Sustainability and Ethics in Supply Chains

Invited Session

MSOM: Sustainable Operations

Chair: Dwaipayan Roy, University of Virginia Darden School of Business, Charlottesville, VA, United States

Co-Chair: Somya Singhvi, USC Marshall School of Business, Los Angeles, CA, United States

1 - Labeling the Supply Chain: Signaling Supply Chain Transparency to Consumers Through Eco-Labels

Katelyn Thompson, University of South Carolina, Columbia, SC, United States, Priyank Arora, Blair Flicker

Increasingly, consumers' buying decisions are influenced by the perceived sustainability of *how products are made*—especially firm efforts to pay workers fair wages and reduce their CO_2 emissions. As such, many firms want to make their supply chains more transparent. Though firms engage in a wide array of sustainability practices, a primary mechanism to disclose this information to consumers, particularly during the buying process, is through a sticker – i.e., an eco-label – printed on product packaging (e.g., Fair Trade, Carbon Reduction). Mapping the complexity of production practices onto a small set of labels, which are often binary, results in information loss and contributes to "label confusion". Several recent studies find that consumers are unclear about the standards that must be met to earn each label and the amount of scrutiny involved in verifying compliance.

Given the widespread use of eco-labels in practice, we undertake a behavioral study to evaluate firms' sustainability practices and labeling strategies, in tandem. We ask: Do firms need separate eco-labels for societal and environmental impact dimensions? Or does information from one dimension spill over onto the other? We consider the following three labeling strategies: (i) augment a single eco-label with precise, quantitative information about firm performance; (ii) complement the existing eco-label with a second eco-label related to the other dimension or; (iii) complement the existing eco-label with a textual self-disclosed claim related to the other dimension. These choices help test the extent of transferability of trust gained via firm's disclosure on one dimension to the other.

2 - Are Chief Sustainability Officers Guardians of Environmental Justice? An Empirical Evaluation

Finn Petersen, University of Minnesota, Carlson School of Management, Minneapolis, MN, United States, Vibhuti Dhingra, Rachna Shah

Environmentally hazardous manufacturing facilities are disproportionately located in underserved communities, which exposes low-income and non-white populations to significantly higher toxic releases and adverse health risks. We study whether a firm governance reform—specifically, the appointment of a Chief Sustainability Officer (CSO)—can mitigate this environmental injustice. Using data on the toxic releases of U.S. manufacturing facilities from 2000 to 2020, we find that appointing a CSO decreases toxic releases by up to 19% and that the effect is particularly pronounced for facilities in underserved communities. We show that this effect is driven, at least in part, by increased implementation of source reduction activities by the treated facilities. Our work sheds new light on how corporate governance reforms act as a control mechanism for addressing environmental injustice, an issue that has proven challenging to regulate effectively.

3 - State Mandated Employment Certificate for Minors Reduces Child Labor Violations in the US

Ashish Kabra, University of Maryland-College Park, College Park, MD, United States, Jiacong Bao

As several states in the US debate loosening child labor laws, we assess the efficacy of existing work permit regulations for child labors. Our analysis reveals that mandating work permits lowers child labor violations, yet their influence is heterogeneous across different industries

4 - Turning a Blind Eye: Buyer-Supplier Relationships and Drug Diversion in the Prescription Opioid Supply Chain

Iman Attari, Indiana University Bloomington, Bloomington, IN, United States, Jonathan Helm, Robert Wiedmer

The opioid crisis has caused severe public health impacts across the United States, leading to approximately 69,000 fatalities in 2020 alone. Academic evidence suggests that the crisis, particularly in its early stages, was fueled by the pharmaceutical industry, with pharmacies playing a significant role through the excessive dispensing of prescription opioid drugs, despite clear indicators of drugs being diverted to non-medical channels. The prevalence of illicit behavior across the supply chain suggests that the current Drug Enforcement Administration (DEA) monitoring strategies, which mandate suppliers to monitor their customers and report suspicious activities, are not effectively curbing the diversion of drugs. Utilizing a comprehensive dataset that captures all opioid drug shipments in the US from 2006 to 2014 and employing a novel approach to detect drug diversion across the supply chain, this study examines how the dynamics of supplier-pharmacy relationships facilitate illicit behavior within the supply network. Our findings reveal the complex mechanisms underlying drug diversion and provide insights that can help governmental agencies, including the DEA, in developing more targeted regulatory measures that address the root causes of illicit behavior more effectively.

WA55

Summit - 447

Power Systems Planning and Operation II: Cutting Edge Strategies

Invited Session

ENRE: Other Energy

Chair: Patricia Hidalgo-Gonzalez, University of California, San Diego, La Jolla, CA, United States

1 - Climate Change and Its Influence on Water Systems Increases the Cost of Electricity System Decarbonization

Patricia Hidalgo-Gonzalez, University of California, San Diego, La Jolla, CA, United States, Julia Szinai, David Yates, Pedro Sánchez-Pérez, Martin Staadecker, Daniel Kammen, Andrew Jones

The electricity sector faces a dual challenge: decarbonization and adaptation to climate change. In many regions, this challenge is complicated by interdependence of electricity and water systems, through hydropower and energy-intensive water resources. By coupling detailed water and electricity system models, we evaluate how climate change alters pathways to carbon-free generation across the Western Interconnect, emphasizing water interactions. We find that grid planning ignoring climate and water linkages misestimates the magnitude, type, and location of capacity needed for decarbonization. By 2050, electricity use could grow by up to 2% annually but up to 8% in July from cooling and water-related electricity demand, while hydropower generation could decrease annually by 23%. Here, we show that to adapt, the region would need to build up to 139 GW of capacity between 2030 and 2050, equivalent to nearly thrice California's peak demand, and could incur up to $150 billion (+7%) in extra costs.

2 - Quantifying the Impact of Energy System Model Resolution on Siting, Cost, Reliability, and Emissions

Anna Jacobson, Princeton University, Princeton, NJ, United States, Denise Mauzerall, Jesse Jenkins

Energy systems models are critical for power sector decision support and decarbonization. Because they are typically written using formulations whose computational requirements increase non-linearly with resolution, large and detailed models are generally intractable even on modern hardware. Scaling dynamics require modelers to omit information for the sake of runtime, affecting siting, cost, and emission results to an unknown degree. We explore the interplay between resolution and results in order to afford decisionmakers greater confidence in recommendation quality. In order to better quantify the impact of coarseness on results, we rely on recent algorithmic innovations via Benders Decomposition to model at higher resolution, and use newly tractable highly granular systems as baselines in comparison with more efficient, low resolution simulations. We find high resolution energy systems models recommend investments with more realistic siting, lower emissions, greater reliability, and less scarcity pricing than analogous models at coarse resolution. We find spatial resolution to be more impactful than temporal, but that errors in model outputs are introduced by the lowest between the two. We see no diminishing returns in model accuracy for several key metrics when increasing resolution. We recommend that modelers use modern mathematical techniques to maximize system granularity and subsequently allocate resources without neglecting any aspect (spatial, temporal, operational) of resolution.

3 - Offshore Wind and Wave Energy Can Reduce Total Installed Capacity Required in Zero Emissions Grids

Natalia Gonzalez, University of California San Diego, La Jolla, CA, United States, Paul Serna-Torre, Pedro Sánchez-Pérez, Ryan Davidson, Bryan Murray, Martin Staadecker, Julia Szinai, Rachel Wei, Daniel Kammen, Deborah Sunter, Patricia Hidalgo-Gonzalez

As the world races to decarbonize power systems to mitigate climate change, the body of research analyzing paths to zero emissions electricity grids has substantially grown. Although studies typically include commercially available technologies, few of them consider offshore wind and wave energy as contenders in future zero-emissions grids. Here, we model with high geographic resolution both offshore wind and wave energy as independent technologies with the possibility of collocation in a power system capacity expansion model of the Western Interconnection with zero emissions by 2050. In this work, we identify cost targets for offshore wind and wave energy to become cost effective, calculate a 17\% reduction in total installed capacity by 2050 when offshore wind and wave energy are fully deployed, and show how curtailment, generation, and transmission change as offshore wind and wave energy deployment increase.

4 - A Nodal Capacity Expansion Model and Data for California Climate Impacts Analyses

Jean-Paul Watson, Lawrence Livermore National Laboratory, Livermore, CA, United States

We consider a power grid capacity expansion optimization model that co-optimizes generation, transmission, and storage given a set of representative weather days. Leveraging the recently introduced CATS synthetic-but-realistic power flow model, we introduce a capacity expansion data set for the state of California in the US, which includes: renewables and load time-series, land use constraints, and requisite unit performance and cost characteristics. We consider both present-day climatology (as represented by the US National Weather Service HRRR data set) and future climatology (as represented by a high-resolution variant of the US Department of Energy's E3SM climate model). Instances for the time periods 2015-2019 and 2040-2045 are analyzed, with solutions obtained in tractable run times via use of advanced progressive hedging decomposition strategies and high-performance computing platforms. Identified expansion plans will be compared and contrast, focusing on impacts of climate change on solution structure.

5 - Hydrological Price Formation in High Renewable Production Cost Modeling Scenarios

Kaleb Smith, Cornell University, Ithaca, NY, United States, Brent Eldridge, Konstantinos Oikonomou

The Western Interconnection system for electricity in the western half of the United States is in a period of market transition given the rapid integration of new variable renewable energy (VRE) capacity and that a large fraction of the region remains vertically integrated. Of particular interest to planners are market systems dominated by zero-marginal cost (ZMC) resources, because current planning scale production cost models (PCMs) may not adequately model the interactions of long-term storage, flexible demand, and high penetration of VRE in the price formation process. Due to the link between the value of storage and potential price volatility shown in these PCMs, accurately modeling the value of hydropower storage in the Columbia River Basin will be important in reliability studies in the Western Interconnection. This work develops a series of operational models for hydropower operation in a simplified grid network representing the Western Interconnection with simulations of four different cases of VRE penetration and demand flexibility. The operational model formulations include an energy budget proxy model, storage equivalent model, fixed head reservoir model, and a convex envelope model. A number of non-linearities are relaxed to reduce computational complexity. Additionally, we simulate on both annual and weekly time-scales, fitting storage values and state of charge values from the annual model as targets of the weekly model.

WA56

Summit - 448

Recent Advances in Stochastic Optimization

Contributed Session

Chair: Yu Zhao, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, 162-0801, Japan

1 - Cognitive Analysis of multiple wind turbine faults using Vision Transformer on Augmented SCADA data

Chufan Wu, City University of Hong Kong, Hong Kong, Hong Kong, Zijun Zhang

This talk presents a novel cognitive-based method for detecting wind turbine component faults, attempting to approach fault detection from a new perspective by leveraging visual clues. In the proposed method, we first transform the SCADA data of target turbine components into images of size 128*128, gathering numerical data of multiple timestamps into informative visual patterns. As the gathering of numerical data decreases the number of trainable faulty samples, we further propose a color adaptation algorithm to integrate with deep convolutional generative adversarial networks (DCGAN) for generating a set of augmented images that show realistic faulty operation patterns. Based on both real-world and augmented SCADA images, a vision transformer is applied to classify images with faulty turbine operation patterns from the normal ones, realizing a high-precision, generalizable, and real-time identification of wind turbine faults. Compared with classic SOTA fault detection methods, we show that, in detecting blade breakage, loosening bolt, and gearbox overheat faults, the proposed cognitive fault detection method achieves a significant reduction in false alarm rates by at least 61.3% while maintaining a correct alarm rate of nearly 100%, serving as a practical solution to AI-assisted wind turbine fault detection.

2 - Automating ML Workflow Orchestration: Strategies for Autonomous Post-Deployment Model Updates

SeongHyun Seo, Sungkyunkwan University, Seoul, Korea, Republic of, Dong-Joon Lim

This presentation is focused on autonomous model update strategies aimed at improving the performance and stability of machine learning (ML) models after deployment. Unlike traditional approaches that often involve periodic retraining or retraining triggered by performance falling below predetermined thresholds, these autonomous strategies make informed decisions regarding the optimal timing and procedures for iterative model training, ensuring long-term sustainability. We propose an adaptive framework to facilitate concept drift detection and adaptation, leveraging data/feature extraction, model management, hyperparameter adjustments, and artifacts versioning. An empirical case study from the Korean steel industry highlights the necessity of autonomous retraining strategies after model deployment and the importance of considering these strategies during model development.

3 - Multi-stage stochastic hub and spoke location-allocation problem

KAZI WAHADUL HASAN, Arizona State University, Tempe, AZ, United States, Pitu Mirchandani

Multi-stage network optimization is a particular area of operations research that frequently emerges in problems related to location-allocation, telecommunication, scheduling, and production systems. In this talk, we will present a multi-stage hub-and-spoke recycling network that considers the single allocation of the spokes. We incorporate two different waste streams, named single stream and source-separated stream, as origins while end markets are destinations. The hub serves as the consolidation point between the origin and destination. Materials from single waste streams remain unsorted and need to go through a material recovery facility (MRF). Our study views the recyclable wastes recovered by the MRF as uncertain due to contamination. We develop a two-stage stochastic mixed integer programming (MIP) model and apply a decomposition approach to solve it. Our decomposition approach is strengthened with some cuts as valid inequalities. We conduct numerical experiments and present our findings to demonstrate the effectiveness of the decomposition approach.

4 - Optimal Control Using Time-series Forecasting Models and Integer Programming

Vinicius Lima, IBM Research, Yorktown Heights, NY, United States, Dzung Phan, Sumanta Mukherjee, Pankaj Dayama, Jayant Kalagnanam

Our talk addresses process control for physical systems with generic forecasting models. We introduce a multi-step set point recommendation framework for industrial process optimization. This method employs multivariate time-series forecasting models to predict long-term trajectories for target variables that must remain within desired operating ranges. We construct a state-based action-response model by simulating various action trajectories and using the response from a global complex forecasting model as the target trajectory. To solve the optimal control problem efficiently, we reformulate it as a mixed-integer linear program (MILP), leveraging MILP representations for action response models. Through experimental evaluations, we demonstrate the efficacy of our proposed techniques.

5 - Evaluating Efficiency in DEA with Consideration of Probabilistic Variations in Data

Yu Zhao, Tokyo University of Science, Shinjuku-ku, Tokyo, Japan

In Data Envelopment Analysis (DEA), both input-oriented and output-oriented models are widely used in practical analysis. Previous studies have applied a bootstrap method to enhance statistical analysis in DEA. In this study, we consider the probabilistic variations present in input-output vectors and propose a tree-based sampling procedure with different orientations for handling the statistical properties of efficiencies. To obtain possible probability distributions of the data, we classify the observed decision-making units into several different clusters and maximize the information gain of each cluster using a Gaussian-based entropy function. The proposed approach shares several characteristics with the bootstrap DEA method. An illustrative example and numerical simulations will be presented.

WA57

Summit - Terrace Suite 1

Advanced Analytics for Healthcare Operations

Invited Session

Health Applications Society

Chair: Gabriel Zayas-Caban, University of Wisconsin Madison, Madison, WI, United States

Co-Chair: Fernando Acosta-Perez, University of Wisconsin-Madison, Madison, WI, United States

1 - Hospital Discharge Prediction Models in Multisite Healthcare Systems

Fernando Acosta-Perez, University of Wisconsin-Madison, Madison, WI, United States, Justin Boutilier, Gabriel Zayas-Caban, Sabrina Adelaine, Frank Liao, Brian Patterson

Care coordination and planning is a complex task for hospital systems, best illustrated by the emergency department (ED) admission process. When deciding where to admit a patient from the ED into the inpatient units, the decision maker needs to balance multiple competing objectives such as the patient waiting time, level of care or specialty fit, and bed blocking of incoming patients. Coordination has become an increasingly complicated task as the result of the increase in hospital mergers and acquisitions, where patients may now have the option of being admitted or transferred to a hospital different from the one that they initially arrived. The use of discharge prediction models has been proposed before as a tool to enhance decision making in the admission process. In multisite healthcare systems, hospitals within the network may interact with each other, and in this case models to predict discharges in one location may benefit from using features/data from other hospitals in the system. In the study we present in this talk, we used more than 170,000 discharge records from two facilities in our partner health system to develop models to dynamically (i.e., near real-time) predict the number of discharges in the next one to four hours in the general care units and explore different strategies to account for the interactions between the different hospitals in the network.

2 - Hospital Occupancy Levels and Inpatient Falls

Jared Chiu, University of Toronto, Toronto, ON, Canada, Vahid Sarhangian, Sarah Tosoni, Laura Pozzobon, Lucas Chartier

Inpatient falls are the highest reported safety events in hospitals. In order to prevent falls, patients typically receive fall risk assessments soon after admission, followed by appropriate interventions. We present semi-Markov models of hospital stay to examine the impact of unit occupancy levels on (1) time-to-fall before receiving an intervention or discharge; and (2) time-to-assessment before fall or discharge. We estimate our models using data from a network of hospitals in Toronto, Ontario and identify a tipping point in unit occupancy level which results in significant reduction in time-to-assessment and increase in risk of fall.

3 - Predicting Long-Term Opioid use via Interpretable Machine Learning

Jingyuan Hu, UCLA Anderson School of Management, Los Angeles, CA, United States, Fernanda Bravo, Elisa Long

Long-term opioid use is associated with increased risks of misuse, addiction, overdose, and death. Early identification of patients at risk provides an opportunity for intervention. This work introduces the Long-Term Opioid Use Risk (LTOUR) model, a simple and interpretable risk-scoring tool developed using an optimization-based machine learning approach. Trained on California's prescription monitoring data, LTOUR performs comparably to state-of-the-art machine learning models while offering greater simplicity and ease of use for healthcare providers, making it a practical tool for real-world applications.

4 - Extended Cna Training Hours and Its Impact on Nursing Homes

Catherine Wang, Wisconsin School of Business, UW-Madison, Madison, WI, United States, Bob Batt, Hessam Bavafa

The study investigates the impact of extended Certified Nurse Aide (CNA) training hours on nursing home quality and staffing. CNAs play a critical role in providing care for residents in nursing homes, yet concerns persist regarding the adequacy of their training. While federal regulations mandate a minimum of 75 training hours, composed of clinical and didactic components, several states have opted to increase this requirement. However, the causal relationship between extended training hours and nursing home outcomes remains unclear. By analyzing changes in training requirements between 2009 and 2019, this study employs a Diff-in-Diff analysis to evaluate the effects on nursing home quality and staffing. The findings address challenges in quality improvement and staffing shortages faced by nursing homes, thereby enhancing resident care and ensuring workforce sustainability within the healthcare system.

5 - Operating theater planning and scheduling for cardiac surgery

Christopher Sun, University of Ottawa, Ottawa, AB, Canada, Thierry Mesana, Amirhossein Moosavi, Bahar Motamedivafa, Jonathan Patrick

Operating theaters are vital yet expensive assets in healthcare facilities, comprising crucial resources like operating rooms, surgeons, intensive care unit beds, and ward beds. Managing these resources efficiently is essential and intricate. This ongoing research investigates the operating theater planning and scheduling at the Ottawa Heart Institute, Canada. We aim to develop a methodology to construct a master surgical schedule and optimize the allocation of elective and emergency patients. The objective functions may include patient wait time, resource utilization and surgery cancellation costs. In this talk, we delve into our problem and solution approach, and share our initial findings.

WA58

Summit - Terrace Suite 2

OR Applications in Healthcare Delivery Systems

Invited Session

Health Applications Society

Chair: Esma Gel, University of Nebraska-Lincoln, Lincoln, NE, 68507, United States

Co-Chair: Thomas Kingsley, Mayo Clinic, Rochester, MN, United States

1 - Predicative Versus Prescriptive Analytics: a 30 Day Hospital Census Forecast for Resource Optimization of Hospital Practice as a Case Study

Thomas Kingsley, Mayo Clinic, Rochester, MN, United States

Our team developed a predictive model that could forecast hospital census across hospital service lines at a large academic hospital. The goal was resource optimization. However, our team discovered resource optimization could not occur in practice with predictive analytics alone. A prescriptive analytics approach was needed to achieve resource optimization goals.

2 - Rationing by Waiting TIME for Outpatient Clinical Appointment Workflows via Novel Patient Access Workflows

Esma Gel, University of Nebraska-Lincoln, Lincoln, NE, United States, Derya Kilinc, Mustafa Sir, Kalyan Pasupathy

Increasing demand for healthcare has necessitated the need to ration care resources by prioritization. We present the use of time windows for rationing by waiting time in outpatient settings. We demonstrate effectiveness for different rationing schemes using real-life transactional data from a leading academic medical center.

3 - Results from Randomized Controlled Trial of an AI Tool for ED and Hospital Operational Efficiency

Alex Ryu, Mayo Clinic, Rochester, MN, United States, Shant Ayanian

Hospital and ED overcrowding remains a significant concern. We designed and implemented an AI model that predicts ED patients' need for hospital admission, with the goal of improving hospital and emergency department efficiency. We discuss the results of a randomized, controlled trial of the AI tool to evaluate performance.

WA59

Summit - Ballroom 1

Strategies for Climate Mitigation and Sustainability

Contributed Session

Chair: Cara Marcy, EIA, Washington, DC, United States

1 - Charting the Climate: Estimating and forecasting greenhouse gas emissions in New Haven county in Connecticut

Peiyao Zhao, University of Massachusetts Amherst, Amherst, MA, United States, Jimi Oke

Climate change has been a pressing issue in society, leading to phenomena such as wildfires, rising sea level, and habitat loss. This is fueled by increasing concentrations of Greenhouse gases (GHG) in the atmosphere. To understand the sources of these emissions and devise effective mitigation strategies, analyzing current emissions pattern and predicting future emissions trends and shifts are paramount. While the development of a nationwide emissions inventory is essential, a regional-level emissions database is significant for local government to design tailored strategies for emissions mitigation. The study presents a greenhouse gas inventory estimation and projection framework specifically tailored for the local area, using New Haven county, Connecticut as a case study. We found that emissions in New Haven totaled 6.97 million metric tons of carbon dioxide in 2021. Among all the emissions from seven sectors (transportation, electricity consumption, solid waste, stationary combustion, agriculture and wastewater), fossil fuel related activities, such as on-road transportation and residential heating, were the primary source of emissions. Due to the fact that we have the most granular data in transportation sector and this sector was the highest emitting contributor, we utilized Gaussian process regression to forecast future vehicle miles traveled (VMT) based on historical trends. Subsequently, we estimated emissions under various VMT scenarios and evaluated the impact of different kernel functions on model performance. Besides, we discover the influence of external factors population, GDP to the model performance. Finally, the policy implication about the carbon dioxide reduction will be provided.

2 - The next generation of long-term energy modeling at EIA

Cara Marcy, EIA, Washington, DC, United States

This presentation will discuss updates related to Project BlueSky. Project BlueSky is an initiative out of the U.S. Energy Information Administration (EIA) to develop a next generation energy systems model, which will eventually be used to produce the Annual Energy Outlook (AEO) and International Energy Outlook (IEO). The new model will be designed to address the complexity and interconnectivity of the modern world and the uncertainty associated with markets, technologies, and international trade.

The EIA is a U.S. statistical agency that collects, analyzes, and disseminates independent and impartial energy information. A core aspect of our mission is to develop multi-decadal, long-term projections of the energy system. Because the energy system is moving at a faster rate than available data, the next generation model will account for future uncertainty in our long-term projections. It will also be nimble so that it can be updated quickly and flexible to handle rapid changes and emerging issues. Our next generation model will be made open source and accessible to the public.

3 - Flexibility-aware co-optimization of integrated low-carbon energy systems for enhancing networks' resilience in extreme events

Yu Weng, MIT, Cambridge, MA, United States, Ruaridh Macdonald

Global climate change induces more frequent extreme events worldwide. Events such as heatwaves, droughts, wildfires, storms, and higher temperatures have posed significant threats to infrastructure safety and the reliable supply of essential services like electricity. According to government data analysis, the occurrence of extreme weather-induced power blackouts has doubled in the past two decades, leading to substantial economic losses and endangering lives, especially during summer and winter. Facing potential disorder, a preventive scheme emphasizing regional self-sufficiency and flexibility analysis is indispensable to ensure the basic service of energy networks under possible stress scenarios. However, there lack of a joint decision-making model that plans for the future where extreme events occur more often. Thus, we propose to a Dolphyn-based framework to optimize the multiple energy carriers for a self-sufficient scheme planning. This planning maintains the basic service inside each region only requiring minimum or affordable extra support by making the most of distributed energy resources. Dolphyn is a decision optimization model enabling long-term low-carbon planning for multiple energy systems, i.e., power, hydrogen, biofuels, liquid fuels, and natural gas. It is an effective toolbox facilitating the network expansion co-optimization planning in the next decades and it can be extended easily. The availability of energy resources is quantified by the flexibility of energy carriers inside each zone. Embedding such flexibility with Dolphyn, the temporal and spatial shifting capability from resources in different energy domains can be utilized to enhance networks' resilience in extreme events.

4 - The Environmental and Economic Impacts of Carbon Tax Policies: A Case Study of GHG Emissions in South Asian Region Countries

Muhammad Suleman, Xi'an Jiaotong University, Xi'an, China, People's Republic of

Climate change is growing as an important issue worldwide, having the potential to have far reaching and irreversible environmental, social, and economic implications. One of the main causes of climate change is emissions of anthropogenic greenhouse gases (GHG). Specifically, CO2 emissions from numerous industrial processes and the combustion of fossil fuels. The power sector contributes significantly to Green House Gas emissions in Pakistan, India, Bangladesh, and other South Asian region countries. The corresponding eliminated for SAR countries would be 11%. If a carbon tax is implemented on SAR countries with higher rates of GHG emissions, hydropower and other renewable energy resources will be used to replace fossil fuels. As a result, power sector GHG emissions in Bangladesh, India, Pakistan, and other South Asia Region countries will be reduced by 55%, 20%, 26%, and 25% from baseline in 2050, respectively. Integrated Energy Model (IEM) in this paper was created using The Integrated Markal-Efom System (TIMES) modeling framework. The TIMES model is a strong tool for analyzing energy scenarios and assessing the impact of prospective policy changes.These findings highlight the significance of reducing GHG emissions through the imposition of a carbon tax.

Key words: Environment, GHG emission, carbon tax, SAR, Hydropower, power sector, IEM, TIMES

WA60

Summit - Ballroom 2

Innovative Approaches for Dynamic Data Analysis and Control

Contributed Session

Chair: James Bailey, Rensselaer Polytechnic Institute, Troy, NY, 12180, United States

1 - Uncertainty Quantification by Set Membership: Convergence Rate Analysis for Control Dynamical Systems

Yingying Li, UIUC, Urbana, IL, United States, Jing Yu, Lauren Conger, Taylan Kargin, Adam Wierman

This paper studies uncertainty set estimation for unknown linear systems. Uncertainty sets are crucial for the quality of robust control since they directly influence the conservativeness of the control design. Departing from the confidence region analysis of least squares estimation, this paper focuses on set membership estimation (SME). Though good numerical performances have attracted applications of SME in the control literature, the non-asymptotic convergence rate of SME for linear systems remains an open question. This paper provides the first convergence rate bounds for SME and discusses variations of SME under relaxed assumptions. We also provide numerical results demonstrating SME's practical promise.

2 - Adaptive Online Learning with LSTM Networks for Energy Price Prediction

Salih Salihoglu, University of Miami, Coral Gables, FL, United States, Ramin Moghaddass

Accurate prediction of electricity prices is crucial for stakeholders in the energy market, particularly for grid operators, energy producers, and consumers. This study focuses on developing a predictive model leveraging Long Short-Term Memory (LSTM) networks to forecast day-ahead electricity prices in the California energy market. The model incorporates a variety of features, including historical price data, weather conditions, and the energy generation mix. A novel custom loss function that integrates Mean Absolute Error (MAE), Jensen-Shannon Divergence (JSD), and a smoothness penalty is introduced to enhance the prediction accuracy and interpretability. Additionally, an online learning approach is implemented to allow the model to adapt to new data incrementally, ensuring continuous relevance and accuracy. The results demonstrate that the custom loss function can improve the model's performance, aligning predicted prices more closely with actual values, particularly during peak intervals. Also, the online learning model outperforms other models by effectively incorporating real-time data, resulting in lower prediction error and variability. The inclusion of the energy generation mix further enhances the model's predictive capabilities, highlighting the importance of comprehensive feature integration. This research provides a robust framework for electricity price forecasting, offering valuable insights and tools for better decision-making in dynamic electricity markets.

3 - Disrupting Labor Exploitation

James Bailey, Rensselaer Polytechnic Institute, Troy, NY, United States, Bahar Cavdar, Yanling Chang

In work settings dominated by short-term contracts and high worker turnover, there is a heightened risk of wage exploitation. In this paper, we introduce a game-theoretical model to study wage theft in such settings. We use the principal-agent framework, categorizing the employer as the principal and the worker as the agent. We analyze the problem when the worker has no awareness of wage theft, and the only deterrence for exploitation is through external inspection and penalties. We show that the problem can be reduced to a single-decision variable and partially characterize the optimal solution to the principal-agent problem to understand the structure of verbal contracts and the resulting wage theft as a function of system parameters. Our analysis extends to different system factors, such as worker skill level, reservation utility, the frequency of inspections, and penalties. This analysis highlights the factors that amplify worker vulnerability, offering strategic insights to mitigate wage theft.

4 - Don't Leave Me on Read! How Avoidance-Based vs. Busyness-Based Attribution of Read Receipts Influences Relationship Outcomes

Yige Wan, Boston University, McKinney, TX, United States, David Fang

In 2012 instant messaging (IM) services such as Facebook Messenger and Whatsapp released the read receipt feature, allowing users to see when someone has read their message. Through a series of 4 online studies (N = 1421) this paper investigated whether individuals would provide a “desire to avoid” attribution or “too busy to respond” attribution when they send a message that was read without reply (RWR) or when they received a message that they RWR and how interpersonal relationships (e.g.trust, closeness, and reliability) become affected by RWR messages. Studies 1 and 2 found that both senders and receivers were more likely to attribute RWR messages to busyness (vs. desire to avoid). Study 3 posited that perceptions of interpersonal factors such as trust, reliability and closeness all dropped for both senders and receivers of RWR messages and that the drop was moderated by an individual’s desire to avoid attribution level. Study 4 further examined this effect in three relationship scenarios of varying closeness levels (parents, friends, and strangers) and found that strangers who sent a message that was RWR tended to underestimate the other party’s desire to avoid them whilst experiencing greater decreases in trust, closeness, and reliability. Additionally, a survey on layman beliefs showed that most participants favor the read receipt function, with nearly 30% being indifferent. When considering reasons why people might not reply after reading a message, the majority believed justifiable reasons such as busyness or forgetfulness, with only a small percentage thinking that they were intentionally ignored.

WA61

Summit - Ballroom 3

AI and Game Theory Applications

Contributed Session

Chair: Guoqing Zhang, University of Windsor, 401 Sunset Avenue, Windsor, ON, N9B 3P4, Canada

1 - A Machine Learning Based on Row Generation Algorithm to Solve the Cooperative Game

Miao Chang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang, Shengnan Zhao

Cooperative game theory constitutes a significant category within the field of game theory. It has demonstrated considerable practical value across various domains due to its facilitation of cooperative endeavors among participants, leading to cost savings. In practical applications, while pure optimization-based solution methods, such as the row generation algorithm, can yield optimal solutions within a sufficient timeframe, they often require a substantial amount of time. Additionally, an important characteristic in real-world scenarios is that the intervals between successive problem-solving instances of cooperative games may be short, and the variations in the input data are typically minor. In this study, we introduce a machine learning-based technique that leverages the solution information from previously solved cooperative game problems to significantly enhance computational performance in solving similar problems in the future. The integration of machine learning technology with mathematical optimization accelerates the speed of traditional row generation algorithms in solving cooperative game problems while still ensuring optimality. As a versatile algorithm, the proposed method is applicable to almost all cooperative game problem. We have tested the performance of the proposed algorithm on several classic and significant cooperative game problems. The experimental results show that the proposed algorithm has a conspicuous performance improvement.

2 - Best of Both Worlds Guarantees for Smoothed Online Convex Optimization

Neelkamal Bhuyan, Georgia Institute of Technology, Atlanta, GA, United States, Debankur Mukherjee, Adam Wierman

We study the smoothed online convex optimization (SOCO) problem where, at each round t, a player plays an action x_t in response to a convex hitting cost and an additional l_2-norm cost for switching actions. This problem class has strong connections to a wide range of application domains including smart grid management, adaptive control, and data center management, where switching-efficient algorithms are highly sought after. Historically, these problems have been studied only from a worst-case viewpoint. We study the SOCO problem in both adversarial and stochastic settings, and in this process, perform the first stochastic analysis of this class of problems. We provide the online optimal algorithm when the minimizers of the hitting cost function evolve as a general stochastic process, which, for the case of martingale process, takes the form of a *distribution-agnostic dynamic interpolation algorithm*, which we call Lazy Adaptive Interpolation (LAI). In this setting, that is stochastic, we prove an $\Omega(T)$ expected regret for the adversarial optimal algorithm in the literature (ROBD). Further, we prove the poor worst-case performance of the stochastic optimal LAI through its sub-optimal competitive ratio. This reveals a trade-off between the stochastic and the adversarial worlds. Finally, we present a best-of-both-worlds algorithm that obtains the *optimal competitive ratio* while simultaneously achieving a *constant regret*.

3 - Configuring a distributed manufacturing supply network with an AI based integrated method

Guoqing Zhang, University of Windsor, Windsor, ON, Canada, Behrang Bootaki

In this talk, we present a supply network problem designed for a distributed manufacturing platform, where the manufacturing facilities are distributed across various locations with the support of Additive Manufacturing (AM) technologies. The proposed problem includes three different types of decisions: location-allocation, production planning, and product delivery routing, named the Location-Production-Routing (LPR) problem. To solve this complex problem, we develop a novel solution method that integrates machine learning surrogate models with a genetic algorithm. Our numerical experiments show that the proposed method can achieve near-optimal solutions. We also provide a lower bound analysis.

WA62

Summit - Signature Room

Online Marketplaces and Incentives

Invited Session

Auctions and Market Design

Chair: Chara Podimata, MIT, Boston, MA, United States

Co-Chair: Keegan Harris, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, United States

Co-Chair: Keegan Harris, CMU, Pittsburgh, PA, United States

1 - User Strategization and Trustworthy Algorithms

Sarah Cen, Massachusetts Institute of Technology, Cambridge, MA, United States, Andrew Ilyas, Aleksander Madry

Many human-facing algorithms---including those that power recommender systems or hiring decision tools---are trained on data provided by their users. The developers of these algorithms commonly adopt the assumption that the data generating process is exogenous: that is, how a user reacts to a given prompt (e.g., a recommendation or hiring suggestion) depends on the prompt and not on the algorithm that generated it. For example, the assumption that a person's behavior follows a ground-truth distribution is an exogeneity assumption. In practice, when algorithms interact with humans, this assumption rarely holds because users can be strategic. Recent studies document, for example, TikTok users changing their scrolling behavior after learning that TikTok uses it to curate their feed, and Uber drivers changing how they accept and cancel rides in response to changes in Uber's algorithm. Our work studies the implications of this strategic behavior by modeling the interactions between a user and their data-driven platform as a repeated, two-player game. We first find that user strategization can actually help platforms in the short term. We then show that it corrupts platforms' data and ultimately hurts their ability to make counterfactual decisions. We connect this phenomenon to user trust, and show that designing trustworthy algorithms can go hand in hand with accurate estimation. Finally, we provide a formalization of trustworthiness that inspires potential interventions.

2 - Matching of Users and Creators in Two-Sided Markets with Departures

Liang Lyu, Massachusetts Institute of Technology, Cambridge, MA, United States, Daniel Huttenlocher, Hannah Li, Asuman Ozdaglar, James Siderius

Many online platforms of today, including social media sites, are two-sided markets bridging content creators and users. Most of the existing literature on platform recommendation algorithms largely focuses on user preferences and decisions, and does not simultaneously address creator incentives. We propose a model of content recommendation that explicitly focuses on the dynamics of user-content matching, with the novel property that both users and creators may leave the platform permanently if they do not experience sufficient engagement. In our model, each player decides to participate at each time step based on utilities derived from the current match: users based on alignment of the recommended content with their preferences, and creators based on their audience size. We show that a user-centric greedy algorithm that does not consider creator departures can result in arbitrarily poor total engagement, relative to an algorithm that maximizes total engagement while accounting for two-sided departures. Moreover, in stark contrast to the case where only users or only creators leave the platform, we prove that with two-sided departures, approximating maximum total engagement within any constant factor is NP-hard. We present two practical algorithms, one with performance guarantees under mild assumptions on user preferences, and another that tends to outperform algorithms that ignore two-sided departures in practice.

3 - [Incentivize Mechanism and Communication Protocol for Federated Bandit Learning] Chuanhao Li

CHUANHAO LI, Yale University, New Haven, CT, United States, Zhepei Wei, Haifeng Xu, Hongning Wang

Federated bandit learning is an extension of the traditional multi-armed bandit problem applied to a distributed setting where multiple agents collaboratively explore a shared environment while preserving data privacy. This approach is critical in scenarios where communication between agents is costly, i.e., a communication protocol balancing regret and communication efficiency becomes essential.

However, most existing works take it for granted that all clients are altruistic about sharing their data with the server for the collective good whenever needed. Despite their compelling theoretical guarantee on regret and communication efficiency, this assumption is overly idealistic and oftentimes violated in practice, especially when the algorithm is operated over self-interested clients, who are reluctant to share data without explicit benefits, or even untruthfully report their cost to get higher payment from the platform. Negligence of such self-interested behaviors can significantly affect the learning efficiency and even the practical operability of federated bandit learning. In light of this, we aim to spark new insights into this under-explored research area by formally introducing an incentivized communication problem for federated bandits, where the server shall motivate clients to share their local data and truthfully report their associated costs. The proposed approach still guarantees the sub-linear regret and communication cost as prior works that forbid agents' strategic behaviors without any overhead, which demonstrates the possibility of simultaneously achieving incentive compatibility and nearly optimal regret in federated bandit learning.

4 - Online Strategic Classification

Saba Ahmadi, TTIC, Chicago, IL, United States, Avrim Blum, Kunhe Yang

In this talk, I discuss an online learning problem in the presence of strategic behavior. In this setting, a sequence of agents arrive one by one and must be classified as positive or negative just like in classic online learning. However, there is a twist: the agents wish to be classified as positive (e.g., consider loan applicants or college applicants) and have the ability to modify their features to a limited extent or at some cost. In particular, agents observe the current prediction rule and will manipulate their features to get classified as positive if they can do so for a cost less than their value for being classified as positive. Agents can manipulate their features in predefined ways, modeled by a manipulation graph. We show how to achieve a bounded number of mistakes when the target function belongs to a known hypothesis set and give upper and lower bounds showing how the graph structure impacts the achievable mistake bounds.

5 - Exploiting Observation Bias to Improve Matrix Completion

Yassir Jedra, MIT, Cambridge, MA, United States

We consider a variant of matrix completion where entries are revealed in a biased manner, adopting a model akin to that introduced by Ma & Chen (2019). Instead of treating this observation bias as a disadvantage, as is typically the case, the goal is to exploit the shared information between the bias and the outcome of interest to improve predictions. To that aim, we consider a natural and simple model where the observation pattern and outcome of interest are driven by the same set of underlying latent or unobserved factors. Under such setting, we propose Mask Nearest Neighbor (MNN), a matrix completion algorithm proceeding in two stages: first, recover (distances between) the latent factors by utilizing matrix estimation for the fully observed noisy binary matrix, corresponding to the observation pattern; second, utilize the recovered latent factors as features and sparsely observed noisy outcomes as labels to perform non-parametric supervised learning. The proposed approach is shown to enjoy entry-wise finite-sample error rates that are competitive with corresponding supervised learning parametric rates. Despite not having access to the latent factors, MNN enjoys such competitive performance by only exploiting the shared information between the bias and outcomes. Finally, through empirical evaluation using a real-world dataset, we find that with MNN, the estimates have 28x smaller mean squared error compared to traditional matrix completion methods, suggesting the utility of the model and the method proposed in this work.

WA63

Regency - 601

AI Development and Business Applications

Invited Session

Information Systems

Chair: Yue (Katherine) Feng, Hong Kong Polytechnic University, Hong Kong

1 - The Impact of AI Summaries on Video Consumption: Insights from a Randomized Field Experiment

Rui QIN, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Yue Feng, Dai Yao

Generative artificial intelligence (AI) technology has been increasingly adopted for content generation in online platforms. A prominent application of this technology is to generate condensed and synthetic summaries of content, which is originally of greater lengths and complexity. To understand the impact of AI summaries on the consumption of original content, we conduct a randomized field experiment on a major video content platform in China. We insert AI summaries to the comments of videos in the treatment group covering a broader range of life cycles, and track various user consumption behaviors of these videos and comparable control videos. Our results demonstrate that the inclusion of AI summaries effectively facilitates video consumption behaviors in diverse aspects, including views, likes, shares, and comments. Further analyses reveal the mechanisms behind these effects, which are contingent on the characteristics of video content, as well as the informativeness and credibility of the summary content. Our findings provide valuable insights into the role played by AI-generated content in online platforms and contribute to the growing body of knowledge surrounding the utilization of AI in content consumption. These insights carry significant implications for marketers, content creators, and platform providers.

2 - Bike-Sharing Systems&Rsquo; Spillover Effects in Local Communities: An Analysis of Business Foot Traffic

Zenan Zhou, Arizona State University, Tempe, AZ, United States, Tian Lu, Elliot Rabinovich

Bike-sharing systems have become an important micro-mobility alternative for many individuals, promoting both local transportation and healthy travel options. As operators strive to maximize user adoption and ensure a broad availability of bikes during station location planning, the establishment of new bike stations may also influence the local economy by affecting foot traffic to nearby businesses. This study seeks to empirically assess the impact of new bike stations on local business footfall. Given that a bike station functions both as a point of departure and a destination, and users may travel to other destinations or return to their starting stations, the effects on nearby business visits are not clearly understood. Addressing these questions could unveil broader benefits of bike-sharing systems and provide valuable insights for operators, local regulators, and business owners. Through an empirical analysis in the Los Angeles metro area, we estimate the effects on local businesses attributed to expansions in the city's bike sharing system and the trips these expansions generate. Based on this estimation, we also offer policy recommendations to help operators, local communities, and city governments maximize the benefits of bike-sharing systems.

3 - Does a Face Matter when It Comes to User Voice? Evidence from a Music Social Media Platform

Jordan He, Lehigh University, Bethlehem, PA, United States, Jia Gao, Jinqi Men, Oliver Yao

AI beauty filters have been increasingly become a standard add-on feature in social media applications, especially with photos and videos. While beauty filters are widely used across various applications and by users of all age groups, there are few studies that explore the effect of the beauty filter on the user-generated content (UGC). Drawing on the self-presentation and social exchange theory, we develop econometric models to fill this literature gap by empirically studying the effect of beauty filters adoption on user engagement such as video likes. We also examine whether the user face matters on a vocal-dominant social media platform that focuses on music entertainment. Interestingly, we find that beauty filters significantly enhance video likes more than singing quality, suggesting that facial appearance is more influential even on this vocal-dominant platform. We further explore interaction effects of singing quality and real facial features (e.g., smiling, eye contact) on the effectiveness of beauty filters. Results show that while beauty filters boost likes, their impact is moderated by both the quality of singing and genuine facial expressions. In contrast, accessories like glasses and headwear amplify the beauty filter effect, despite their general negative association with video likes. Implications for theory and practice are discussed.

WA64

Regency - 602

Graphs and Algorithms in Social Media: Theories and Applications

Invited Session

Social Media Analytics

Chair: Pantelis Loupos, University of California, Davis, Davis, CA, United States

1 - Network Formation and Dynamics among Multi-LLMs

Marios Papachristou, Cornell University, Ithaca, NY, United States, Yuan Yuan

Social networks shape opinions, behaviors, and information dissemination in human societies. As large language models (LLMs) increasingly integrate into social and professional environments, understanding their behavior within the context of social interactions and networks becomes essential. Our study analyzes LLMs' network formation behavior to examine whether the dynamics of multiple LLMs are similar to or different from human social dynamics. We observe that LLMs exhibit key social network principles, including preferential attachment, triadic closure, homophily, community structure, and the small-world phenomenon when asked about their preferences in network formation. We also investigate LLMs' decision-making based on real-world networks, revealing that triadic closure and homophily have a stronger influence than preferential attachment and that LLMs perform well in network formation predictions. Overall, our study opens up new possibilities for using LLMs in network science research and helps develop socially aware LLMs by shedding light on their network formation behaviors and exploring their impacts on social dynamics.

2 - Graph Neural Networks for Causal Inference Under Network Confounding

Pantelis Loupos, University of California, Davis, Davis, CA, United States

This paper studies causal inference with observational network data. A challenging aspect of this setting is the possibility of interference in both potential outcomes and selection into treatment, for example due to peer effects in either stage. We therefore consider a nonparametric setup in which both stages are reduced forms of simultaneous equations models. This results in high-dimensional network confounding, where the network and covariates of all units constitute sources of selection bias. The literature predominantly assumes that confounding can be summarized by a known, low-dimensional function of these objects, and it is unclear what selection models justify common choices of functions. We show that graph neural networks (GNNs) are well suited to adjust for high-dimensional network confounding. We establish a network analog of approximate sparsity under primitive conditions on interference. This demonstrates that the model has low-dimensional structure that makes estimation feasible and justifies the use of shallow GNN architectures.

WA65

Regency - 603

Applying Artificial Intelligence Models to Text Mining for Leveraging Business Value

Invited Session

Artificial Intelligence

Chair: Nohel Zaman, University of Alabama at Birmingham, Birmingham, AL, United States

1 - Discovering Unintended Consequences of App Usability from Online Reviews of AR/VR-integrated Apps using Text Mining Technique

Nohel Zaman, University of Alabama at Birmingham, Birmingham, AL, United States

This study focuses on applying the supervised text mining approach to extract the key terms specific to unintended consequences of AR-VR app usability. The results may help app developers and product designers improve the feature concerns of AR/VR-integrated apps and develop strategies to address feature suggestions to improve the app usability.

2 - Combining NLP and Data Visualization to Analyze Voluminous Narrative Text: Applications for Journalism, Law, Philosophy, and Research

Michel Mitri, James Madison University, HARRISONBURG, VA, United States

This presentation pertains to a software application called Story Analyzer (SA), which combines natural language processing (NLP) and interactive data visualizations to produce dashboards that help users navigate complex text documents. SA uses Stanford's CoreNLP as its NLP engine, which provides these services: sentence splitting, tokenization, parts-of-speech tagging, lemmatization, named entity recognition, dependency parsing, coreference resolution, and temporal tagging. Story Analyzer uses all these services for its information extraction. After information extraction, SA generates interactive dashboards depicting people, groups, their interactions, places and times of the interactions, and key themes from a story narrative. During my talk, I intend to discuss the CoreNLP's services, how they are used by SA to analyze narrative and discursive text, applications of these to current political and legal issues in the news, merging NLP results with interactive data visualization technology, and technical issues with implementing meaningful text-analytic dashboards for enlightening readers of complex text. I will demo some SA dashboards to illustrate the user experience. I also encourage you to view this video to get an idea of a typical SA dashboard: https://youtu.be/mnnF2md_Lh8.

3 - Detecting ICO Scams with Explainable AI

Ruilin Wang, UMass Lowell, Lowell, MA, United States, Xiaobai Li

Initial coin offerings (ICOs) are an increasingly popular method for startups to raise funds for business development. In this process, companies sell new digital tokens or cryptocurrencies to investors in exchange for capital. These tokens can be traded with other investors or used to purchase the company's goods or services. However, ICO projects are not bound by specific disclosure or audit requirements, leading to an unregulated environment that raises concerns about ICO-related fraud. Prior works have explored features that could indicate potential ICO fraud and developed methods to predict ICO scams. Machine learning and AI techniques have been used to predict ICO scams, but there is often a trade-off between accuracy and explainability. While complex models may offer greater prediction accuracy, they typically lack transparency and are harder to understand. This paper aims to improve the accuracy of ICO fraud prediction by using machine learning and AI techniques to extract features from ICO white papers. At the same time, it also adapts Large Language Models (LLMs) to enhance the interpretability of the results. This research will support legislative and regulatory efforts in the ICO market, help potential investors to identify ICO scams, and contribute to the application of LLMs.

WA66

Regency - 604

Data Integration for Learning and Decision-Making

Invited Session

Artificial Intelligence

Chair: Kaizheng Wang, Columbia University, New York, NY, United States

1 - Non-Stationary Bandits with Auto-Regressive Temporal Dependency

Djallel Bouneffouf, IBM, Poughkeepsie, NY, United States

Traditional multi-armed bandit (MAB) frameworks, predominantly examined under stochastic or adversarial settings, often overlook the temporal dynamics inherent in many real-world applications such as recommendation systems and online advertising. This paper introduces a novel non-stationary MAB framework that captures the temporal structure of these real-world dynamics through an auto-regressive (AR) reward structure. We propose an algorithm that integrates two key mechanisms: (i) an alternation mechanism adept at leveraging temporal dependencies to dynamically balance exploration and exploitation, and (ii) a restarting mechanism designed to discard out-of-date information. Our algorithm achieves a regret upper bound that nearly matches the lower bound, with regret measured against a robust dynamic benchmark. Finally, via a real-world case study on tourism demand prediction, we demonstrate both the efficacy of our algorithm and the broader applicability of our techniques to more complex, rapidly evolving time series.

2 - Transfer learning for contextual multi-armed bandits

Changxiao Cai, University of Michigan, Ann Arbor, MI, United States

We study the problem of transfer learning for nonparametric contextual multi-armed bandits, where we have access to data pre-collected from source bandits to assist decision-making in the target bandit. We explore two cases: (1) the source dataset consists of expert demonstrations in which arm assignments predominantly reflect optimal choices; (2) it is produced by an exploratory policy in which suboptimal arms are uniformly randomly explored. The minimax rates of convergence for the regret are established and novel transfer learning algorithms that provably attain these minimax regrets are proposed. The results quantify the contribution of the data from the source domains for learning in the target domain in the context of nonparametric contextual multi-armed bandits, providing a precise characterization of the dependence on the distribution shift and behavior policy.

3 - Speeding up Policy Simulation in Supply Chain Reinforcement Learning

Tianyi Peng, Columbia University, New York, NY, United States, Vivek Farias, Joren Gijsbrechts, Aryan Khojandi, Andrew Zheng

Simulating a single trajectory of a dynamical system under some state-dependent policy is a core bottleneck in policy optimization algorithms. The many inherently serial policy evaluations that must be performed in a single simulation constitute the bulk of this bottleneck. To wit, in applying policy optimization to supply chain optimization (SCO) problems, simulating a single month of a supply chain can take several hours.

We present an iterative algorithm for policy simulation, which we dub Picard Iteration. This scheme carefully assigns policy evaluation tasks to independent processes. Within an iteration, a single process evaluates the policy only on its assigned tasks while assuming a certain 'cached' evaluation for other tasks; the cache is updated at the end of the iteration. Implemented on GPUs, this scheme admits batched evaluation of the policy on a single trajectory. We prove that the structure afforded by many SCO problems allows convergence in a small number of iterations, independent of the horizon. We demonstrate practical speedups of 400x on large-scale SCO problems even with a single GPU, and also demonstrate practical efficacy in other RL environments.

4 - Sequential Decision Making with Expert Demonstrations under Unobserved Heterogeneity

Keertana Chidambaram, Stanford University, Standford, CA, United States, Vahid Balazadeh, Vasilis Syrgkanis, Viet Nguyen, Rahul Krishnan

We study the problem of online sequential decision-making given auxiliary demonstrations from experts who made their decisions based on unobserved contextual information. These demonstrations can be viewed as solving related but slightly different tasks than what the learner faces. This setting arises in many application domains, such as self-driving cars, healthcare, and finance, where expert demonstrations are made using contextual information, which is not recorded in the data available to the learning agent. We model the problem as a zero-shot meta-reinforcement learning setting with an unknown task distribution and a Bayesian regret minimization objective, where the unobserved tasks are encoded as parameters with an unknown prior. We propose the Experts-as-Priors algorithm (ExPerior), an empirical Bayes approach that utilizes expert data to establish an informative prior distribution over the learner's decision-making problem. This prior enables the application of any Bayesian approach for online decision-making, such as posterior sampling. We demonstrate that our strategy surpasses existing behaviour cloning and online algorithms, as well as online-offline baselines for multi-armed bandits, Markov decision processes (MDPs), and partially observable MDPs, showcasing the broad reach and utility of ExPerior in using expert demonstrations across different decision-making setups.

5 - An Adaptive Neuron Reset Threshold for Maintaining Plasticity in Continual Learning

Adam Jozefiak, Massachusetts Institute of Technology, Cambridge, MA, United States, Vivek Farias

Recent literature has demonstrated that a neural network's ability to adapt to new data progressively deteriorates, commonly referred to as the loss of plasticity. Empirically, we find that many existing techniques for addressing plasticity loss are sensitive to the choice of their corresponding hyperparameter(s) where a single log-scale deviation often results in a substantial degradation of performance.
Therefore, these existing continual learning methods necessitate an exponentially larger hyperparameter search space, creating an additional computational cost to their deployment. To remedy this, we propose the adaptive neuron-reset threshold algorithm (ART) that addresses plasticity loss by selectively resetting neurons according to an adaptive threshold. ART has the flexibility to specify a single network-specific threshold, layer-specific thresholds, or neuron-specific thresholds, tailoring reset-threshold parameters to both the order and architecture of each layer. We verify through extensive experiments, on common continual learning variants of the MNIST and ImageNet datasets, that ART outperforms existing continual learning techniques while simultaneously being robust to its choice of hyperparameters. While ART still retains hyperparameters, they are naturally interpretable and our experiments demonstrate that any "reasonable" choice results in a performant algorithm. We additionally perform an ablation study with a fixed reset-threshold variant of ART, demonstrating that ART's adaptive layer-specific reset-thresholds

contribute to a more performant intervention, beyond simply alleviating the need for an exhaustive hyperparameter search.

WA67

Regency - 605

Innovative Applications in Healthcare and Wellness

Contributed Session

1 - Optimizing Home Care Worker Assignments: A case study in British Columbia

Kibele Sebnem YILDIRIM, The University of British Columbia, Kelowna, BC, Canada, Mohammadsadra Nejati, Amir Ardestani-Jaafari, Hamidreza Eslami

The vital role of home care workers underscores the significance of efficiently managing their assignments, especially in regions where the availability of such workers is limited. The challenge lies in optimizing their schedules and routes to serve a diverse client base spread across various locations. Our aim is to cluster clients effectively, ensuring that each nurse is assigned a manageable workload within a close geographic proximity. By minimizing the driving distance between clients assigned to the same nurse, we strive to enhance operational efficiency and timely service delivery. To achieve this objective, we propose two methods. Firstly, leveraging regression techniques, we generate closed-form predictions of driving distances between clients using data from the Google Maps API. Subsequently, we explore the application of deep learning algorithms to refine these predictions further. Finally, we employ optimization techniques to allocate clients to nurses, considering both the predicted distances and the predetermined threshold for maximum travel distance. We validate our proposed methods through a case study conducted in the province of British Columbia, Canada.

2 - What kind of physicians do patients prefer, live streaming or not? Empirical evidence from online health communities

Binhua Ye, Tongji University, Shanghai, China, People's Republic of, Jiantong Zhang

The emergence of live streaming has given rise to a new economic paradigm, blending with the medical services, resulting in a novel modality known as physician live streaming (PLS). However, there is a dearth of evidence regarding the impact of PLS on physicians' online consultation performance. Drawing on the Interaction Ritual Chain (IRC) theory, this paper aims to address the existing research gap, with a primary focus on the influencing factors of live streaming characteristics, consultation characteristics, and the mediating role of Parasocial Relationship (PSR) between physicians and patients. We extract datasets from online health communities in China, which encompass live streaming and online consultation information of 951 physicians, and construct an quarterly-level unbalanced panel dataset. Employing the fixed effects model and difference-in-differences with propensity score matching (PSM-DID), the results reveal a significant positive effect of PLS on physicians' online consultation performance, with PSR mediating the relationship significantly. Moreover, the Emotional Energy (EE) between the physician and patient emerges as a salient moderating factor in the nexus between the physician's live-streaming service and their online consultation performance.

3 - Dynamic Linear Modeling of Social Determinants of Health During the Covid Pandemic

Rasim Musal, Texas State University, Austin, TX, United States, Tevfik Aktekin, Tahir Ekin

The Covid-19 pandemic has affected every aspect of society. We discuss the changes in the effects of social determinants of health for the duration of 77 biweeks on Covid-19 Mortality in the 58 counties of California. We specifically focus on the changes in the effect of median household income and poverty levels while controlling for other variables, including but not limited to vaccination rates, median age, population size and density as well as spatial effects within the counties. In investigating random and spatial effects we use Besag York Mollie model and utilize STAN platform for the application of the Bayesian framework.

4 - Machine learning for optimal test admission in the presence of resource constraints

Dmitry Krass, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ramy Elitzur, Eyal Zimlichman

Developing rapid tools for early detection of viral infection is crucial for pandemic containment. This is particularly crucial when testing resources are constrained and/or there are significant delays until the test results are available – as was quite common in the early days of Covid-19 pandemic. We show how predictive analytics methods using machine learning algorithms can be combined with optimal pre-test screening mechanisms, greatly increasing test efficiency (i.e., rate of true positives identified per test), as well as to allow doctors to initiate treatment before the test results are available. Our optimal test admission policies account for imperfect accuracy of both the medical test and the model prediction mechanism. We derive the accuracy required for the optimized admission policies to be effective. We also show how our policies can be extended to re-testing high-risk patients, as well as combined with pool testing approaches. We illustrate our techniques by applying them to a large data reported by the Israeli Ministry of Health for RT-PCR tests from March to September 2020. Our results demonstrate that in the context of the Covid-19 pandemic a pre-test probability screening tool with conventional RT-PCR testing could have potentially increased efficiency by several times, compared to random admission control.

WA68

Regency - 606

AI and Analytics in Operations and Logistics

Contributed Session

Chair: Christian Truden, University of Klagenfurt - Department of Operations, Energy, and Environmental Management, Klagenfurt, 9020, Austria

1 - Artificial Intelligence (AI) Security for Logistic Disruption Mitigation in Distributed Uav Swarms and Other Autonomous Cyber-Physical Systems

Lida Hagh, Old Dominion University, Norfolk, VA, United States, Mohammad Ghasemigol

New cyber-physical systems (CPSs) that integrate the physical and digital (cyber) spaces must be protected on critical infrastructures such as the Port of Virginia. Among new CPSs to support port operations, Unmanned Aerial Vehicles (UAV) are increasingly used in monitoring and communications for remote reconnaissance missions, surveillance operations, and supporting command and control. Many of these applications are distributed in nature, requiring coordination, planning, and often runtime reconfiguration to conduct operations. Traditionally, human decision-making controls UAV operations' movement and task completion. However, the literature is extensive indicating the limitations of this approach. Artificial Intelligence (AI) has emerged as a critical tool to overcome these limitations. We propose that methods would make risk mitigation and informed decision-making challenging. Unlike traditional intrusion detection systems, ML-based systems require less human intervention and are more effective in detecting new attacks.

2 - Apache Spark-Based Approach for Automated Polygon Generation in Maritime Operation Areas

Gabriel Fuentes, Norwegian School of Economics, Bergen, Norway, Justin McGurk, InKyung Choi, Kristoffer Overholdt Solberg, Eric Glenjen

Statistics generation from Automatic Identification System (AIS) data has been beneficial for maritime research. The process from raw data to statistics generation usually involves finding vessel port visits with AIS data and polygons outlining berths and anchorage areas. However, manual production of detailed polygons restricts such studies to local or regional contexts. To mitigate this, we present an automated approach for generating polygons that can identify critical maritime operational areas, including berths, anchorage zones, and Single Buoy Mooring areas. While current methods offer general approximations of port areas based on serial algorithms with limited AIS data, our technique extends this work by enabling recognition of specific berths and anchorages. This method, designed to leverage the power of parallel computing, can process worldwide AIS data within a big data architecture, such as Apache Spark.

3 - Disasters, Infrastructure, and Supply Chains

Seth Guikema, University of Michigan, Ann Arbor, MI, United States

Natural hazards regularly impact infrastructure systems, and both infrastructure outages and the hazard itself impacts supply chains. The dependence of supply chains on interdependent infrastructure poses significant analysis challenges for supply chain operators. This talk presents new approaches for estimating (1) loss of interdependent infrastructure services during natural hazards and (2) loss of functionality of supply chains due to both direct hazard impacts and loss of infrastructure services. The modeling approach is a combination of stochastic simulation and network theory and uses only publicly-available information to estimate infrastructure loss of service and supply chain impacts.

4 - ML-Based Prediction of Vehicle Fleet Performance in Routing Applications

Christian Truden, University of Klagenfurt - Department of Operations, Energy, and Environmental Management, Klagenfurt, Austria, Mike Hewitt

We present a method to perform a comprehensive analysis of the fleet composition problem that is suitable for most variants of the vehicle routing problem. Its basic principle is to estimate a fleet's performance by using the company's delivery planning tools in a black-box fashion. In a case study, we analyze the fleet size and mix for a fictional grocery home delivery service. A fleet comprising multi-compartment vehicles is employed, where each compartment is designated for storing groceries at specific temperature zones tailored to their storage requirements. In general, the stakeholders are interested in finding a fleet configuration that enables good performance regarding defined key performance indicators (KPIs). Seasonal demand changes occur in nearly all types of routing applications. Therefore, we aim to identify fleet configurations that ensure consistent and satisfactory performance across all seasons. We do not propose a methodology for choosing a fleet. This is because stakeholders may consider multiple KPIs when making fleet composition decisions, and these KPIs may be conflicting and vary by scenario. Thus, we focus on a method for predicting the values of multiple KPIs for a given fleet.

WA69

Regency - 607

AI-Driven Cybersecurity

Invited Session

Artificial Intelligence

Chair: Tung Cu, Northeastern Illinois University, Chicago, IL, United States

1 - AI-Enabled Search Engines: Ethics, Privacy and Security

Tung Cu, Northeastern Illinois University, Chicago, IL, United States

The study conducted a systematic review of the social and technical challenges of AI-enabled Search Engines related to ethics, privacy, and cybersecurity. Firstly, the ethical dimensions of AI search engines are profound, as these tools shape our perception of the world but are often manipulated by commercial interests. The ethical challenges include issues like the distribution of knowledge, global dominance, and decision-making processes in displaying results. Secondly, privacy concerns are a central theme in the discourse surrounding AI. The use of AI for user profiling and data manipulation has raised alarms about the need for stringent data protection measures. Past research reflects public apprehension about the use of personal data by AI systems, especially in sensitive areas like education. They emphasize the importance of user awareness and acceptance of data tracking and profiling by AI systems. Finally, cybersecurity is intricately linked with AI advancements. Others address the security implications of AI, noting that ethical AI utilization in cybersecurity and hacking requires transparency, accountability, and fairness. The rapid advancement of Generative AI models like ChatGPT and Google Bard further complicates this landscape, as these tools can be used both defensively and offensively, raising new security and privacy concerns. Inclusion, we proposed a list of recommendations and policies to tackle the challenges. The body of research underscores the need for an interdisciplinary approach to address the ethical, privacy, and security challenges posed by AI search engines. This approach must balance technological innovation with the protection of individual rights and societal values.

2 - Exploratory Study on Factors Affecting Believability of Video-Based Misinformation

Calvin Cheng, Hong Kong Polytechnic University, Hong Kong , Hong Kong, Vanessa Liu

User-generated videos are made easy and quick by Generative AI image-to-video and text-to-video technologies. The extant literature on misinformation focuses primarily on texts or images. This study aims to fill this gap by adopting multiple theoretical perspectives to examine the factors affecting believability of video-based misinformation on social media. Drawing on the concepts of Autonomous Sensory Meridian Response, we test the effect of video-specific artifacts such as sound quality in shaping believability. While prior studies show that believability leads to viral behavior such as reposting, we investigate the reciprocal effect of virality on believability based on the social influence notion. Leveraging the social presence theory, we compare believability in AI generated videos vis-à-vis regular videos. An experiment will be conducted with 700 participants to empirically verify our hypotheses. The findings should be of theoretical and practical contributions to both the domains of social media and cyberlaw.

WA70

Regency - 701

Electric Reliability and Resilience

Invited Session

ENRE: Electricity

Chair: Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, 78712, United States

1 - The Value of Coordination for Restoring Power and Wireless Networks

Rachel Moglen, The University of Texas at Austin, Austin, TX, United States, Benjamin Leibowicz, Alexis Kwasinksi

In the aftermath of natural disasters, the rapid restoration of critical infrastructures—in this paper we focus specifically on power and wireless cellular networks—is imperative, yet challenging due to the complexities of their interactions. Cellular networks rely on the power network for electricity to operate, meaning their restoration processes cannot be siloed from power restoration. To support optimal power and cellular network restoration, we develop a framework that prescribes actions that maximize demand served by both networks. We analyze coordination dynamics between power network decision makers and wireless carriers, ranging from perfect coordination—where both power and cellular restoration efforts are optimized simultaneously—to zero coordination—where decision makers make restoration plans and execute independently and without knowledge of the other decision makers' restoration plans. We apply this framework to a case study in Guayama, Puerto Rico following a 50-year design storm hurricane and thereby explore how coordination can impact restoration outcomes in practice. We find that perfect restoration coordination can significantly improve cellular restoration outcomes at a limited cost to power restoration outcomes, relative to zero coordination. We also explore the impacts of partial coordination, where wireless carriers identify a limited number of critical assets that are prioritized in the power network's restoration problem. We find that this partial coordination can capture some of the benefits of perfect coordination relative to zero coordination, when carefully executed. Finally, we find that information sharing without modification to power restoration has limited ability to improve cellular restoration outcomes over the zero coordination case.

2 - The Short-Term Strategic Response of Residential Customers to Long-Duration Power Interruptions

Juan Pablo Carvallo, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Climate change is driving more frequent and intense extreme weather events, which in turn are producing longer and widespread electric service interruptions. These interruptions can have substantial economic impacts, but quantifying this impact remains elusive and understudied. This paper studies the response of residential customers to long duration power interruptions by identifying their actions after the event and the economic costs of both these actions and the interruption. We develop a novel survey instrument and apply it to over 1000 residential customers in the ComEd Illinois service territory for fictional summer and winter 1-, 3-, and 14-day long interruptions. Analyzing the response using probit regression, we find that customers are significantly more likely to relocate if they do not own a generator, if they live in multi-family buildings, and if they are poor. Rural customers, in contrast, are 27% less likely to relocate compared to their urban counterparts, revealing better adaptation to the hardship brought by lack of electricity service. We estimate that, overall, residential customers are 12% more likely to relocate for each additional day interrupted. Economic survey responses reveal that relocation is substantially more expensive than staying home, even when accounting for the direct impacts of the interruption, but over 80% of customers would relocate. Analyzing relocation costs by income bracket reveals that poor customers would spend, on average, their entire monthly income during a 14-day interruption, revealing substantial inequality concerns about resilience response. These results have important implications for the equality of electric interruption impacts, and for

3 - Anticipating Uncertain Infrastructure Interdependencies: A Case Study of Coupled Power-Transportation Networks

Alireza Rangrazjeddi, Vanderbilt , Nashville, TN, United States

Infrastructure systems consist of interdependent relationships, where various components interact through physical, functional, or logical connections. A disruption in one component can have a cascading effect, impacting the functionality of other components. Similarly, resilience planning in one network may impact the overall resilience of the system of coupled infrastructure networks.

Infrastructure interdependencies are often driven by physical connections between these networks represented by connected facilities and equipment. However, under uncertain conditions, such as climate-driven disasters, these interdependencies can become uncertain and change dynamically as the disaster is evolving. For example, in the case of power and transportation networks, power outages can impact traffic conditions due to failure in traffic signals. Consequently, blocked or congested roads can hinder the restoration of power outages due to significant delays and reduced accessibility of repair crews dispatched to restore the power grid. As such, understanding and evaluating the stochastic nature of infrastructure interdependencies is important to better prepare for and respond to disasters.

This study aims to develop a framework for understanding these complex interdependencies between components of infrastructure networks. We propose to use stochastic block models, a class of statistical network models, to evaluate the probability of such interdependencies evolving over the course of a disaster and as failures cascading across multiple network components. The study will specifically focus on the interdependencies between power and transportation networks. This probabilistic approach seeks to enhance the resilience and robustness of critical infrastructure networks against various disruptions.

4 - Resilient Capacity Expansion Planning for Power Grids in a Changing Climate

Berk Sahin, University of Texas-Austin, Austin, TX, United States, Erhan Kutanoglu, Andrea Staid

Extreme weather events have a significant impact on the power grid, affecting the operation of lines, substations, and generators, causing outages and disrupting service to populations that increasingly rely on power. Scientists believe that climate change will exacerbate these events. Therefore, integrating climate change and long-term resilience needs into the grid planning processes is essential for future power grid performance. In this research, we analyze ways to make optimal capacity expansion planning decisions, aiming to meet the future resilience, supply and demand needs of the grid while considering the nature and uncertainty of extreme events and climate change. We use realistic test cases based on actual power grids to compare the benefits of the proposed approaches considering the projected growth in various aspects of the grid including population, renewables, and electrification.

WA71

Regency - 702

Data Analytics in Manufacturing Industry

Invited Session

Data Mining

Chair: Dohyun Kim, Myongji University, Yongin-si, N/A

Co-Chair: Jason Shin, Mississippi State University, 230 Brook Avenue, Starkville, MS, 39759, United States

1 - Development of a Graph Generation Model for Linking Heterogeneous Domain Graphs

SuA Yoon, Myongji University, Yong-In, Korea, Republic of, Dohyun (Norman) Kim

Analyzing heterogeneous data provides comprehensive and innovative insights that cannot be obtained from a single data type, making it essential for the advancement of knowledge and applications. This study develops a deep learning-based algorithm to generate graphs that connect heterogeneous data sources. Specifically, we present a model that integrates expert and corporate graphs to create a new graph linking experts and companies. This approach is expected to enhance connections and collaboration between experts and companies, thereby fostering innovation and improving decision-making processes.

2 - Llm-Based Recommendation System Between Manufacturing Enterprises and Experts

Ki Jeong Choi, Myongji University, Yong-in, Korea, Republic of, Dohyun (Norman) Kim

The increasing complexity of industrial processes and the rising demand for specialized knowledge have underscored the need for recommending experts to address challenges faced by manufacturing enterprises. Traditional recommendation systems struggle with this due to the textual nature of existing data on enterprise information, including challenges and expert profiles. Recent advancements in Large Language Model (LLM) technology have shown promise in overcoming these difficulties. We propose a recommendation system that utilizes LLMs to analyze textual data from manufacturing enterprise information and expert profiles, providing suitable expert recommendations to resolve specific enterprise challenges.

3 - Designing Electric Power Grid Investment Plans for Economic Growth and Energy Security

Kwon Gi Mun, California State Polytechnic University - Pomona, Irvine, CA, United States, Sungyong Choi, Mark Rodgers, Wenbo (Selina) Cai, Gül Çulhan Kumcu

This study presents a Generation Expansion Planning (GEP) model that determines the optimal capacity expansion investment plan while considering economic growth, energy security, and renewable energy considerations. Specifically, the motivation of this study is to leverage the power grid capacity expansion decisions in developing countries to facilitate sustainable economic development. To accomplish this, first, a relationship between energy generation and Gross Domestic Product (GDP) is established with an econometric model. Then, this linear relationship is integrated into the GEP model to support the expansion planning process. According to an exhaustive search of the energy literature, this is the first comprehensive study that endogenously integrates the GDP econometric function into the constraints of the GEP model. This study aims to provide policy development and analysis insights, especially for resource-rich, energy-poor countries. Further, this study presents a roadmap that developing countries can follow to increase GDP while reducing fossil fuel dependency and avoiding CO2 emissions from these sources.

4 - How Does the "Golden Screw" Problem Affect New Order Dynamics

Jason Shin, Mississippi State University, Mississippi State, MS, United States, Jason Miller, Judy Whipple

Industry practitioners and policy makers have a strong interest in understanding dynamics associated with manufacturing firms' new orders. In the United States, many organizations including the Institute for the Supply Management (ISM®), multiple branches of the Federal Reserve Banking System, and the Census Bureau each publish data on new orders to inform practitioners and policy makers. This research explores how the impact of shifts in manufacturing firms' new orders impacts their production, recruitment of workers, and their suppliers' lead times. We draw on economic theory regarding how economic activity behaves during periods of chronic shortage to explain why increases in new orders will have a less pronounced effect on production and employment but a greater impact on supplier lead times during periods of pronounced shortage. To test our theory, we assemble a unique archival dataset that melds ISM's Purchasing Manager Index (PMI) data with the Census Bureau's Quarterly Survey of Plant Capacity Utilization to obtain measures of shortages. We test our hypotheses using time series techniques. We find results consistent with our predictions, which have important implications for theory, practice, and public policy.

WA72

Regency - 703

Applications on Data Analytics

Invited Session

Data Mining

Chair: Jiao Wu, Northern Illinois University, DeKalb, IL, United States

1 - Exploring the Impact of Virtual Instructor Design and Study Activities on Student Engagement and Emotional Responses

Jiao Wu, Northern Illinois University, DeKalb, IL, United States

This work explores students' emotional responses and study engagement with different types of virtual instructors (human-like versus stylized). We propose that the design of virtual instructors and the nature of study activities (lecture-based versus interaction-based) influence students' engagement and emotional responses. Utilizing Emotional Design Theory, we will analyze patterns and correlations in student interactions and feedback to draw meaningful insights. This research could contribute to the design of educational information technology and teaching strategies. The findings may enhance educational outcomes and support student well-being in technology-mediated learning environments.

2 - Enhancing Influencer Marketing: The Impact of Visual Attributes and Influencer–Product Gender Congruence on Consumer Engagement

Ying Wang, Northern Illinois University, DeKalb, IL, United States, Jaeki Song

This study examines the dynamics of influencer marketing, specifically the impact of visual attributes and gender congruence between influencers and products on consumer engagement. Utilizing framing theory and gender congruence literature, we suggest that visual elements like colorfulness, brightness, and vividness play crucial roles in driving active consumer engagement. Our analysis of 2,000 YouTube videos in feminine and masculine product categories indicates that colorfulness and vividness enhance engagement, whereas brightness may decrease it. Gender congruence between influencer and product significantly intensifies the influence of visual attributes on engagement. The findings provide a comprehensive understanding of how visual attributes and gender congruence interact in influencer marketing, offering theoretical insights and practical guidance for optimizing marketing strategies and enhancing the effectiveness of influencer marketing.

3 - Analyzing Key Factors Influencing Online Reviews of Historically Black Colleges and Universities (HBCUs) through Text Mining

Jiaxi Luo, Midwestern State University, Wichita Falls, TX, United States

This study employs text mining techniques to identify key factors influencing online reviews of Historically Black Colleges and Universities (HBCUs) using data from niche college rankings. Sentiment analysis, topic modeling, and keyword extraction algorithms are applied to uncover prevalent themes and sentiments in the reviews. The identified factors are analyzed to determine their impact on the perception of HBCUs. Additionally, a comparative analysis between HBCUs and other universities is conducted to highlight any distinct differences. The findings contribute to a better understanding of the unique challenges and strengths of HBCUs, providing insights for administrators, policymakers, and researchers to enhance the educational experience and support the growth of these vital institutions. The comparative analysis emphasizes the significance of HBCUs in the higher education landscape.

4 - Digital Transformation and Carbon Emissions: a Deep Neural Network Based Latent Dirichlet Allocation Approach

Jiang Hu, Texas A&M University San Antonio, San Antonio, TX, United States, Ying Teng

This study investigates the impact of digital transformation on firms' carbon emission intensity (CEI). Employing a deep neural network (DNN)-based Latent Dirichlet Allocation (LDA) to assess the level of digital transformation and Economic Input-Output Life Cycle Assessment (EIO-LCA) to quantify CEI, we analyzed data based on 1958 firm-year observations from 673 Chinese listed companies between 2010 and 2020. Our findings reveal an inverse U-shaped relationship between digital transformation and CEI. This relationship varies by firm type, specifically between firms operating in environmentally sensitive industries (ESF) and those operating in non-environmentally sensitive industries (NESF). While digital transformation has no significant impact on ESFs' CEI, it has a notable inverse U-shaped relationship with NESFs' CEI. Moreover, certain board characteristics, such as independence, gender diversity, and age diversity, negatively moderate the relationship between digital transformation and CEI. In addition to enrich the existing literature, our study provides valuable insights for government and industry practitioners.

5 - Data Analytics

Peng Li, Rutgers University, Newark, NJ, United States

WA73

Regency - 704

Data Mining Best Paper Competition (General Track)

Award Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

Co-Chair: Andi Wang, University of Wisconsin-Madison, Madison, WI, United States

1 - Trustworthy Feature Importance Avoids Unrestricted Permutations

Emanuele Borgonovo, Bocconi University, Milan, Italy

coming soon

2 - Penalized Additive Gaussian Process for Screening and Optimization of Quantitative and Qualitative Factors in Black-Box Systems

Yongxiang Li, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

coming soon

3 - Triple Component Matrix Factorization: Untangling Global, Local, and Noisy Components

Naichen Shi, University of Michigan, Ann Arbor, MI, United States

In this work, we study the problem of common and unique feature extraction from noisy data. When we have N observation matrices from N different and associated sources corrupted by sparse and potentially gross noise, can we recover the common and unique components from these noisy observations? This is a challenging task as the number of parameters to estimate is approximately thrice the number of observations. Despite the difficulty, we propose an intuitive alternating minimization algorithm called triple component matrix factorization (TCMF) to recover the three components exactly. TCMF is distinguished from existing works in literature thanks to two salient features. First, TCMF is a principled method to separate the three components given noisy observations provably. Second, the bulk of the computation in TCMF can be distributed. On the technical side, we formulate the problem as a constrained nonconvex nonsmooth optimization problem. Despite the intricate nature of the problem, we provide a Taylor series characterization of its solution by solving the corresponding Karush–Kuhn–Tucker conditions. Using this characterization, we can show that the alternating minimization algorithm makes significant progress at each iteration and converges into the ground truth at a linear rate. Numerical experiments in video segmentation and anomaly detection highlight the superior feature extraction abilities of TCMF.

4 - Cost-aware Bayesian optimization via the Pandora's Box Gittins index

Qian Xie, Cornell University, Ithaca, NY, United States

coming soon

WA74

Regency - 705

Decision Support for the Path to Net-Zero Emissions

Invited Session

ENRE: Energy-Climate

Chair: Jesse Jenkins, ZERO Lab, Princeton University, Princeton, NJ, United States

Co-Chair: Edgar Virgüez, Carnegie Science, Stanford, CA, United States

Co-Chair: Jacqueline Dowling, Carnegie Science, 260 Panama St., Providence, 94305, United States

1 - MACRO: A holistic planning tool for net-zero emissions energy and industrial systems

Jesse Jenkins, Princeton University, Princeton, NJ, United States, Filippo Pecci, A.D.T Perera, Luca Bonaldo, Sambuddha Chakrabarti, Ruaridh Macdonald, Jun Wen Law, Dharik Mallapragada

In a transition to net-zero emissions, energy and industrial systems will become increasingly coupled, as complex flows of electricity, hydrogen, bioenergy, and CO2 create strong interconnectedness across multiple energy vectors and industrial supply chains. Ideally, energy and industrial system planning models should be able to co-optimize multiple interconnected energy and industrial supply chains. Discrete planning decisions are also important, in order to capture the significant economies of unit scale that characterize many assets in energy networks and industrial facilities. And with electrification expanding the role of electricity across the economy and weather-dependent renewable resources moving to the center of the energy supply mix, such models should retain the high temporal and spatial resolution found to be critical in prior power systems planning research while considering uncertainty across multiple weather years.

These challenges motivate the development of MACRO, a next-generation holistic planning tool for net-zero emissions energy and industrial systems. This talk will discuss the computational advances necessary to enable capacity expansion models for integrated energy and industrial system planning with expansive scope, granular spatial resolution, high temporal resolution, and discrete planning decisions. That includes harnessing both temporal and spatial decomposition strategies to break apart these large-scale problems and take full advantage of parallel computing capabilities.

2 - Net-zero emissions energy systems

Steven Davis, Stanford University, Stanford, CA, United States

Net emissions of CO_2 by human activities—including not only energy services and industrial production but also land use and agriculture—must approach zero in order to stabilize global mean temperature. A review paper published in 2018 described the technological and economic challenges to deliver all modern energy services with net-zero CO_2 emissions. I will briefly recap the main takeaways of that work, discuss important technological and research progress during the past 6 years, and highlight current research frontiers. As the review paper did, the presentation will address four areas in particular: aviation, long-distance transportation and shipping; structural materials; reliable electricity; and carbon management.

3 - Integrated Energy Systems Perspective on Blue and Green Hydrogen in Net-Zero Energy Systems

Jun Wen Law, MIT Energy Initiative, Massachusetts Institute of Technology, Cambridge, MA, United States, Bryan Mignone, Dharik Mallapragada

The versatility of Hydrogen (H2), with multiple production technologies and potential end use applications, is one factor driving its growing appeal. However, this versatility leads to multiple potential pathways and cross-sectoral interactions that complicate its assessment. Here, we apply an open-source, multi-sector capacity expansion model (MACRO) to investigate the drivers of green and blue H2 deployment in a cost-optimized, net-zero energy system. MACRO represents the spatial and temporal coupling between the power grid, H2 infrastructure (production, storage, and transmission), and spatially resolved CO2 storage and biomass resources. In a case study of the contiguous U.S., we find that H2 deployment is primarily affected by H2 infrastructure such as H2 storage and pipelines. When such resources are unconstrained, green H2 accounts for the largest share of H2 supply, as it exploits the H2 infrastructure to maximize electrolytic H2 production at locations and times of low electricity prices and vice versa. In contrast, blue H2, which requires additional investments in negative emissions technologies and CO2 transport and storage, accounts for the largest share of H2 supply when H2 infrastructure is limited. Counterintuitively, increasing blue H2 operating flexibility with limited H2 storage and pipeline deployment increases green H2 deployment, as it also enables operational flexibility of green H2. Our analysis also sheds light on the value of grid-connected vs. islanded H2 systems under cost-optimized net-zero constraints. We find that grid-connected systems meet a greater share of H2 demand using electrolyzers, highlighting potential synergies of co-optimized electricity and H2 production in decarbonized energy systems.

4 - Improving the Performance of Risk-Adjusted Stochastic Unit Commitment for Clean Electricity

Dimitrios Floros, Duke University, Durham, NC, United States, Mauricio Hernandez, Kyle Bradbury, Dalia Patino-Echeverri

The increased likelihood of extreme weather events and the expansion of variable renewable energy (VRE) sources underscore the need for improved uncertainty characterization in electric power systems. Traditional real-time operation decisions for power generators and storage are predicated on single-point forecasts of electricity demand and energy supply for the planning horizon. Cognizant of the error in such forecasts, electricity system operators schedule assets so that the system has enough operating reserves to account for real-time deviations in demand and renewable energy production, which may lead to elevated reserve targets and system costs under higher uncertainty. A promising alternative that reduces system costs and enhances reliability is the stochastic unit commitment (SUC) methodology. Rather than using a single-point forecast, SUC utilizes an ensemble of plausible forecasts to inform operational decisions by pre-positioning all power-generating and storage assets to satisfy demand under each scenario. The applicability of SUC, however, is contingent upon the quality of the probabilistic forecast ensembles and the ability to solve the problem within the required time constraints. To this end, we introduce the solution learning approach (SLA) that identifies patterns in past SUC solutions and utilizes these patterns to solve future SUC problems promptly. We benchmark our framework in stochastic unit commitment within the Duke Energy system—the largest vertically integrated utility in the United States—showcasing the SLA's potential as a practical and scalable solution for enhancing SUC processes.

WA75

Regency - 706

Energy Justice II: Equity in Energy Systems

Invited Session

ENRE: Other Energy

Chair: Weihang Ren, University of Florida, Gainesville, FL, United States

Co-Chair: Yongpei Guan, University of Florida, Gainesville, FL, United States

1 - Transmission Capacity Planning Through the Lens of Energy Justice

Weihang Ren, University of Florida, Gainesville, FL, United States, Yongpei Guan

In this paper, we present a novel transmission capacity planning framework that integrates energy justice principles and constraints. The transmission capacity directly influences the binding constraints within an ISO's economic dispatch model, ultimately determining energy prices. The proposed model incorporates various energy justice metrics to inform transmission capacity planning, introducing complementary constraints into the economic dispatch model. Our framework identifies optimal transmission capacity investments, prioritizing disadvantaged communities for enhanced energy access and equity.

2 - Distribution Grid Planning With Equity Objectives

Luis Rodriguez-Garcia, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, Miguel Heleno

This talk presents an optimization model for an equity-informed deployment of distributed energy resources in power distribution grids. More specifically, this model decides on the location of solar generation and battery energy storage to enhance the resilience of power distribution systems to heat waves, while ensuring that the distribution energy resource installation is equitably distributed based on community vulnerability to extreme heat. The model considers a metric that captures the impact of power outages during extreme heat as an objective, initially analyzed on individual buildings and then aggregated for a system-level overheating inequity assessment. Through this implementation, state and local authorities can support decisions to foster equitable and resilient energy infrastructure development.

3 - Understanding Energy Burden in New England Households

Jie Xu, George Mason University, Fairfax, VA, United States, Siwei Li, Saba Siddiki, Feng Qiu

We analyzed the distribution of energy burden among households living in New England states using 5-year American Community Survey data. We trained machine learning models to identify major factors associated with energy burden.

WA76

Regency - 707

Industrial Decarbonization II: Optimizing Hydrogen Use to Decarbonize the Transportation and Industrial Sectors

Invited Session

ENRE: Other Energy

Chair: Ahmed Abdulla, Carleton University, Ottawa, ON, K1S 5B6, Canada

Co-Chair: Ahmed Abdulla, Carleton University, Ottawa, ON, Canada

1 - Hydrogen Microgrids Could Facilitate Diesel Exit in Remote and Northern Communities

Ahmed Abdulla, Carleton University, Ottawa, ON, Canada, Ian Maynard

Most remote and northern communities in the Arctic continue to rely on diesel for their electrical and thermal needs, including to serve their industries and livelihoods. Communities and governments are working toward diesel exit strategies, but the role of hydrogen technologies has not been explored. These could serve both electrical and thermal loads, reduce emissions, and enhance energy security and self-determination. Here, we present results from energy optimization models that compare the cost, emissions, and resource requirements of both diesel and hydrogen microgrids across a large sample of varied communities. These models optimize the buildout of each microgrid and track how hydrogen is stored and used to serve both electrical and thermal loads over the course of the year. Our model shows that hydrogen microgrids remain less costly over their lifetimes than diesel systems across the majority of communities studied, though they require large overbuilding of electric power generation capacity to meet both electrical and thermal needs.

2 - Using Direct Air Capture Waste Heat for Hydrogen Production

Kieran Graham, Carleton University, Ottawa, ON, Canada, Kristen Schell, Erick Arwa

Direct air capture (DAC) of carbon dioxide is a promising technology to enable climate change mitigation. The liquid solvent DAC (LSDAC) process is one of the leading technologies being piloted. However, LSDAC uses a high-temperature regeneration process which requires a lot of thermal energy. Although current LSDAC designs incorporate pre-heat cyclones and a heat recovery steam generator to enable heat recovery, these do not maximize the use of the heat in the products of calcination. In this work, a process model is designed to maximize the use of waste heat from the LSDAC process to drive water electrolysis to produce hydrogen. Water use and net energy production from this new process are analyzed. Further examination identifies the price at which hydrogen production is prioritized.

3 - Optimizing Hydrogen Delivery to Heavy Duty Trucks in a Decarbonized Electricity Supply System

Andrew Moffat, Carleton University, Ottawa, ON, Canada

Heavy road freight transportation annually produces 66 million tonnes of greenhouse gas emissions in Canada, representing 9% of the country's total emissions—on par with its entire electricity system. To date, little progress has been made in decarbonizing this subsector. Current efforts in the field of freight decarbonization center around interventions in truck operations. These measures include platooning, autonomous operation, aerodynamic improvements, and electrification. Moreover, this research rarely considers the impacts of freight decarbonization on other sectors of the energy system, including the electricity system. Meanwhile, electricity system models often ignore the technological richness and uncertainties associated with freight decarbonization pathways. This research aims to bridge the gap by analyzing how freight decarbonization pathways affect the electricity system. The integrated assessment model includes: 1) a physics-based model to calculate truck energy use on road segments; 2) an uncertainty-informed demand load projection to 2050 that considers freight technology, electrification of light duty transport, climate, and population growth; and 3) a generation expansion planning model of the Alberta power system that determines where capacity should be built to serve the high demands of freight decarbonization. Results demonstrate that a switch to battery-electric traction could reduce total energy use to 15 PJ, while a transition to hydrogen would result in total energy consumption of 35 PJ (compared to 55 PJ of diesel), with substantial benefits to downstream emissions as both hydrogen and batteries have zero tailpipe emissions.

4 - The Impact of Uncertainty in Low-Carbon Industrial Demand on The Decarbonization of a Canadian Province: Results of a Generation Expansion Planning Model

Caleb Stairs, Carleton University, Ottawa, ON, Canada

As the world deals with the increasingly frequent and severe effects of climate change, it has become imperative that collaboration and proactive measures be taken at local, national, and international levels to reduce greenhouse gas emissions. Actions are required to shift energy use to low carbon and renewable sources. Forecasting models are needed to predict future energy needs and analyze the cost of meeting those needs.

This research analyzes the uncertainty in future energy demand in the small eastern Canadian province of New Brunswick and explores the effect of that uncertainty on generation expansion planning across the province. To date, provincial energy forecasting has been limited and mainly done by industry and government with little transparency. Such efforts often use scenario analysis rather than fully examining the uncertainties inherent in this activity. This research uses Monte Carlo simulation to develop predictions of energy needs integrating population growth, effects of climate, increased electric vehicle use and electrification of industry. A generation expansion planning model reveals the effect of uncertainty in this future demand on the cost and mix of resources required to achieve net-zero emissions by 2050.

WA77

Regency - 708

Simulation & Analytics for Sustainable and Resilient Urban Networks - Session 1

Invited Session

Computing Society

Chair: Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA

1 - On Local Topological Measures and Network Vulnerability Patterns: a Comparative Analysis

Nazanin Tajik, Mississippi State University, Mississippi State, MS, United States, Saviz Saei

This study examines how complex systems are vulnerable to the increasing impacts of climate change and extreme weather events. The focus is on the transportation infrastructure in the United States. The research measures the susceptibility of external events to random, targeted, and localized disruptions. Recent findings on local topological measures are synthesized to illustrate network vulnerability under various conditions. The research utilizes the 35 most commonly used local measures in critical infrastructures to quantify connectivity, accessibility, and criticality levels of (i) 15 popular topologies in social and physical network studies, (ii) the Nguyen and Dupuis network, and (iii) Sioux Falls Network. The results from topological and non-topological analysis demonstrate the effectiveness of local measures in explaining real-world critical infrastructure performance under different disruption scenarios.

2 - Decentralized Scheduling in Network-of-Networks Post-Disruption to Enhance Resilience: a Multi-Agent Reinforcement Learning Study

Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA, United States, Shima Mohebbi

Optimizing task scheduling in interdependent systems characterized as network-of-networks post-disruption is crucial for enhancing resilience. Traditional centralized scheduling models fail to capture the dynamic complexities of these networks governed by interconnected nodes. This study proposes a decentralized scheduling approach using multi-agent reinforcement learning (MARL). The methodology leverages a hybrid simulation model that captures organizational and network evolution dynamics, enabling agents to autonomously carry out scheduling policies via MARL. A novel reward shaping method is incorporated into the reinforcement learning framework, allowing agents to learn from past experiences and schedule actions based on previous scheduling decisions. The increasing frequency of natural disasters underscores the importance of optimal scheduling for restoration efforts in critical infrastructures (CIs). By combining decentralized decision-making with a holistic view of the system, the proposed approach offers a solution that is both comprehensive and adaptable to the intricate nature of network-of-networks. The proposed model is applied to water distribution and mobility networks in Tampa, FL, demonstrating the efficacy of the decentralized approach in restoring components post-disruption and improving network resilience.

3 - Discrete Event Simulation Model for Capturing the Impacts of Virtual Health Adoption Rates for Outpatient Mental Health Clinics

Aaron Sporrer, US Air Force, Wright-Patterson AFB, OH, United States, Robert Curry, Anna Svirsko

The Defense Health Agency (DHA) is in the early stages of a historical advancement in the way healthcare is delivered to its 9.5 million beneficiaries world-wide. With an aim to improve access to care and patient experience with a reduction in resourcing, the agency has introduced a new “digital first” methodology to the Family Practice and Behavioral Health service lines at five venture sites with the intent of expanding to its remaining 703 Military Treatment Facilities in Fiscal Year 2025 and forward. The objective of this study is to assess how various adoption rates for virtual care may impact the resourcing requirements for the Behavioral Health Clinic at one of the venture sites (Wright Patterson Medical Center) and gain insights on how these advances may impact Military Healthcare at-large. To meet this objective, we use a discrete event simulation to model the implementation of virtual care within the Behavioral Health Clinic at Wright Patterson Medical Center. In particular, we study how an increase in virtual care appointments will affect provider utilization rates. Our analysis finds that while the transition to a “digital first” model has a modest impact on the resourcing requirements within the Behavioral Health Clinic, it has a substantial impact on workplace availability for Active Duty patients across the force.

4 - Integrated Intermodal Freight Models and Tools for Efficiency and Resiliency – Use of POLARIS agent-based simulation tool

Jesus Osorio, University of Illinois Urbana-Champaign, Urbana, IL, United States, Natalia Zuniga

The U.S. freight transportation industry is responsible for a significant portion of the nation’s greenhouse gas (GHG) emissions, and its under-capacitated infrastructure has exacerbated its vulnerability to disruptions. One of the primary barriers to the enhancement of freight network decarbonization and resiliency is the lack of an integrated, overarching, and data-driven modeling framework that can be used to predict, optimize, and provide decision-support for intermodal freight logistics operations.

This presentation introduces the proposal for the development of a series of integrated data and computational models, as well as associated executable computer programs, that would enable modeling and decision making on the nation’s intermodal freight logistics systems across water, roadway, and railway modes. System dynamics analysis under disruption scenarios will also be used to assess the system resilience. All these data and model capabilities will be integrated into POLARIS, a high-performance, computationally efficient tool that implements advanced travel and freight demand modeling, dynamic traffic assignment, and transportation simulation into an integrated modeling platform.

WA78

Regency - 709

Optimization and Robust Linear Models

Contributed Session

Chair: Bo Tang, University of Toronto

1 - Uncertainty quantification for iterative algorithms in linear models with application to early stopping

Kai Tan, Rutgers University, Piscataway, NJ, United States, Pierre C Bellec

This paper investigates the iterates $\hbb^1,\dots,\hbb^T$ obtained from iterative algorithms in high-dimensional linear regression problems, in the regime where the feature dimension p is comparable with the sample size n, i.e., $p \asymp n$. The analysis and proposed estimators are applicable to Gradient Descent (GD), proximal GD and their accelerated variants such as Fast Iterative Soft-Thresholding (FISTA). The paper proposes novel estimators for the generalization error of the iterate $\hbb^t$ for any fixed iteration t along the trajectory. These estimators are proved to be $\sqrt n$-consistent under Gaussian designs. Applications to early-stopping are provided: when the generalization error of the iterates is a U-shape function of the iteration t, the estimates allow to select from the data an iteration $\hat t$ that achieves the smallest generalization error along the trajectory. Additionally, we provide a technique for developing debiasing corrections and valid confidence intervals for the components of the true coefficient vector from the iterate $\hbb^t$ at any finite iteration t. Extensive simulations on synthetic data illustrate the theoretical results.

2 - A Learning Based Large Neighborhood Search Framework for Mixed-Integer Linear Programming

Yu Sun, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, 110819, China, People's Republic of, Ying Meng, Lixin Tang

Mixed-Integer Linear Programming (MILP) is one of the most crucial modeling approaches in the field of combinatorial optimization. Since most combinatorial optimization problems are NP-hard, solving large-scale MILP models is time-consuming. In this paper, we propose a learning-based large neighborhood search framework to effectively solve the general MILPs. In the proposed framework, we first adopt the hypergraph partitioning technique to decompose the model into an arrow corner block structure, and then employ Graph Neural Networks (GNN) to identify integer variables in the model that are likely to appear in the optimal solution. Based on the obtained corner block structure and the information provided by the GNN, a construction heuristic algorithm named Decomposition Kernel Search (DKS) and an improved heuristic algorithm named Decomposition Graph Neighborhood Search (DGNS) are designed. Experiments on the MIPLIB2017 dataset demonstrate that our framework substantially improves the solving efficiency of MILP solvers.

3 - Sensitivity analysis for linear change of the constraint matrix of an LP

Quentin Louveaux, University of Liège, Liège, Belgium, Bardhyl Miftari, Damien Ernst, Guillaume Derval

Understanding the variation of an optimal value with respect to change in the data is an old problem of linear optimization. In this paper, we focus on a linear problem in which a matrix D modifies linearly the coefficients of the constraint matrix A. It can be used to analyze the impact of multiple linear changes in the constraint matrix. As a new approach to the problem, we derive five bounds which allow us to avoid recomputing the problem for numerous changes in the matrix A and provide guarantees on the objective function’s behavior. We discuss an iterative algorithm that use these bounds to compute an approximation to the original function within a given error threshold. We analyze the performance of the bounds on toy and real-world problems and demonstrate the effectiveness of the approach.

4 - Data-Driven Distributionally Robust Chance-Constrained Linear Matrix Inequalities

Fengjie Liang, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, Jianqiang Cheng, Kai Pan

This paper presents approximation and reformulation techniques for Distributionally Robust Chance-Constrained Linear Matrix Inequalities (DRCCLMIs) to overcome the computational challenges arising from multidimensional integration and non-convexity of feasible sets. DRCCLMIs seek a robust solution which ensures the satisfaction of chance constraints across a wide range of possible distributions within an ambiguity set. In particular, we consider a data-driven ambiguity set which contains all potential distributions that share the same moment information. We first derive an inner approximation for the general form of DRCCLMIs to cope with common constraint structures encountered in the real world. The key method we use is the Conditional Value-at-Risk (CVaR) approach, which allows us to approximate DRCCLMIs in a way that ensures a certain solution quality while maintaining robustness. In addition, we also derive exact reformulations for a special case, i.e., DRCCLMIs with block matrix structure embedded in the linear matrix inequalities. Notably, our approximation and reformulation techniques help transform the original DRCCLMIs into more tractable semidefinite programming (SDP) problems, thereby simplifying the computational process and improving the accuracy of the solution. The effectiveness of these techniques are demonstrated through numerical studies on two real-world applications: truss topology design problem and calibration problem.

5 - CaVE: A Cone-Aligned Approach for Fast Predict-then-optimize with Binary Linear Programs

Bo Tang, University of Toronto, Toronto, ON, Canada, Elias Khalil

The end-to-end predict-then-optimize framework, also known as decision-focused learning, has gained popularity for its ability to integrate optimization into the training procedure of machine learning models that predict the unknown cost (objective function) coefficients of optimization problems from contextual instance information. Naturally, most of the problems of interest in this space can be cast as integer linear programs. In this work, we focus on binary linear programs (BLPs) and propose a new end-to-end training method to predict-then-optimize. Our method, Cone-aligned Vector Estimation (CaVE), aligns the predicted cost vectors with the normal cone corresponding to the true optimal solution of a training instance. When the predicted cost vector lies inside the cone, the optimal solution to the linear relaxation of the binary problem is optimal. This alignment not only produces decision-aware learning models, but also dramatically reduces training time as it circumvents the need to solve BLPs to compute a loss function with its gradients. Experiments across multiple datasets show that our method exhibits a favorable trade-off between training time and solution quality, particularly with large-scale optimization problems such as vehicle routing, a hard BLP that has yet to benefit from predict-then-optimize methods in the literature due to its difficulty.

Wednesday, October 23, 9:30 AM - 10:45 AM

WB01

Summit - 320

Integrating AI in Healthcare Operations to Enhance Service Delivery

Invited Session

Service Science

Chair: Zhenzhen Jia, Xi'an Jiaotong University, Xi'an, N/A

1 - AI's Assistance in Disease Diagnosis: When Patients Cannot Access Advanced Tests

Zhenzhen Jia, Xi'an Jiaotong University, Shaanxi, China, People's Republic of, Jianqiang Hu, Kejia Hu, Sriram Venkataraman, Qingchen Wang, Ning Zhang

The accuracy of diagnostic processes significantly influences patient outcomes and carries considerable implications for service efficiency and quality within healthcare systems. However, this accuracy can often be compromised by financial, psychological, or physical barriers that restrict access to advanced diagnostic tests, especially among rural patients. In response to these constraints, our study explores the deployment of artificial intelligence to maintain high diagnostic accuracy, regardless of the availability of advanced test results. This paper introduces SIP-HD, a novel deep learning model that performs simultaneous imputation and prediction on high-dimensional data. Notably, our model utilizes the imputation process to enhance the accuracy of predictions. When benchmarked against various two-step models, one-step models, and doctors' preliminary diagnoses, SIP-HD exhibited superior performance in terms of accuracy, area under the curve, and F1 score. The proposed model offers significant potential to enhance diagnostic decision-making in real-world scenarios, particularly when advanced tests are inaccessible. Its adaptability amplifies its value, allowing it to serve as both a supportive diagnostic tool and a telemedicine aid, with a potential extension to an online self-diagnosis platform. These applications hold promise for improving service efficiency and quality within healthcare while also offering substantial cost savings.

2 - Nonparametric Data-Driven Algorithms with Inventory Competition and Demand Correlation

Lei Lei, Chongqing University, Chongqing, China, People's Republic of

We study stochastic inventory planning of two competitive firms with lost sales and instantaneous replenishment under inventory competition, i.e., a proportion of consumers will switch to another retailer in the case of stockout. The primary demands are allowed to be correlated, but the knowledge of the joint distribution of the primary demands is not available. At the same time, we observe only the sales quantity in each period without knowing the lost sales. Furthermore, it is impossible to distinguish the primary and spillover demands for each firm. We develop an algorithm to make the ordering decision in each period based only on historical sales data, and conduct numerical experiments to demonstrate the effectiveness of our proposed algorithm.

3 - Signaling quality with return policies: a social learning perspective

Huiwen Guo, Chongqing University, Chongqing, China, People's Republic of, Bingsheng Liu, Yinghua Shen, Yuan Chen

Amid the increasingly prevalent online shopping, consumers often exhibit considerable uncertainty in their product valuation. In response, firms may provide lenient return policies (e.g., partial returns or full returns) to encourage consumers to make purchase because these policies not only stimulate demand but may also signal quality. Yet, benefits derived from lenient return policies to firms may not be that straightforward, considering the increasing returned products and the signaling cost (as a high refund). Situation becomes more complicated when consumers can utilize online reviews, which often serve as a free channel for quality disclosure, to update their belief of product valuation. Through a two-period stylized model, this research analyzes the informational roles of return policies when product quality (either low or high) is the firm’s private information, as well as the strategic interaction between the firm and social learning consumers. We especially focus on the effectiveness of return policies as the signal of product quality, and how online reviews interact with it. Our analysis yields three main findings. First, the optimal partial-return policies serve as an effective signal of high quality for products with small salvage value or high exogenous price, where separating equilibrium exists. Second, for most of the firms, offering partial-return policies is more advantageous for them than no-return or full-return policies considering the increased firm profit. Third, the synergy between return policies and online reviews can better reveal the quality of products. Intuitively, by being distinguished from low-quality firms easily, high-quality firms benefit more from online reviews.

4 - Rethinking Appointment Scheduling: Addressing Counselor Burnout and Improving Patient Outcomes

Yuchen Liang, National University of Singapore, Singapore, Singapore

Since the growing demand for mental health services has placed significant strain on psychological counsellors, posing unique challenges for counselling clinic operations, we study the impact of mental stress on counsellor performance and clinic operations within the appointment scheduling setting. Utilizing a data-set from a counselling clinic in a Singaporean university, we examine the effects of workload and task complexity on counsellor behaviour. To improve clinic operations, we propose a scheduling heuristic that integrates workload and task complexity considerations, and develop a discrete-event simulation model to compare the results.

WB02

Summit - 321

Modeling and Online Monitoring Techniques with Health Applications

Invited Session

Quality, Statistics and Reliability

Chair: Lu You, University of South Florida, Lutz, FL, 33620, United States

1 - Reduced-Rank Cox Regression for Analyzing High-Dimensional Multi-Response Survival Data with Hidden Confounders

Kai Yang, Medical College of Wisconsin, Milwaukee, WI, United States

In this talk, we present a low-rank Cox model that correlates multiple survival responses with high-dimensional mixed-type covariates in the presence of unmeasured confounders. In confounded regression, recent advancements focus on generalized or multivariate linear models, typically estimating latent components through linear factor models and drawing inference based on debiased Lasso. In many applications, however, the existence of discrete covariates makes linear factor models unsuitable. To overcome this limitation, we quantify the association between covariates and confounders using a generalized linear factor model and propose an efficient parallel algorithm for recovery of hidden confounders. Moreover, we design an effective estimation scheme to consistently determine ranks of coefficient matrices and compute estimators of regression parameters in the multi-response Cox model, employing the proposed adaptive nuclear norm and sparse-group Lasso penalty. To facilitate inference, we develop an innovative approach named hidden conditional independence regression, which transforms high-dimensional inference into low-dimensional regression problems. This idea is general and potentially can be extended to other high-dimensional inference problems involving unmeasured confounders. Our method outperforms existing approaches such as the debiased Lasso. It accommodates rare-event responses and between-response correlations. Theoretical and numerical studies demonstrate its effectiveness in practice.

2 - Online Risk Monitoring of Multiple Diseases by a Dynamic Screening System

Zibo Tian, University of Florida - Department of Biostatistics, Gainesville, FL, United States, Peihua Qiu

Disease early detection and prevention (DEDP) is an important topic in medical and public health research. Because disease risk factors are usually observed sequentially over time, DEDP is a sequential decision-making problem and statistical process control (SPC) charts turn out to be a powerful tool after the major complexities of the observed data (e.g., time-varying distributions and serial correlation) are properly addressed. In the literature, several SPC charts have been developed for solving the DEDP problem, but they are designed for detecting a single disease. In practice, however, we are often concerned about multiple diseases (e.g., different cardiovascular diseases), and there are no existing SPC methods designed for detecting multiple diseases yet due to its complexity. In this paper, a new dynamic screening system (DySS) is proposed for detecting multiple diseases. The new method first quantifies a patient's risk to each disease in concern at the current observation time, and then compares the quantified risk pattern with the regular risk pattern of non-diseased people that is estimated from a training dataset by a flexible longitudinal data modelling approach. The cumulative difference between the two risk patterns by the current observation time is used for determining whether a given patient has any of the multiple diseases in concern. Numerical studies show that the proposed method works well in different scenarios.

3 - Online Monitoring and Early Detection of Influenza Incidence Rates Using Exponentially Weighted Spatial Lasso: A Case Study in China During 2015-2020

Yuhang Zhou, University of Florida - Department of Biostatistics, Gainesville, FL, United States, Shoumi Sarkar, Yang Yang, Peihua Qiu

The People's Republic of China, since 2000, has developed an extensive nationwide surveillance network, focusing on various diseases, including influenza. This network's data is characterized by significant spatial and temporal variations, presenting a complex challenge in the real-time detection of deviations in the spatio-temporal (ST) process, such as outbreaks. The data structure in this network is intricate, encompassing complex ST variations and correlations. One notable aspect is that process shifts, often signaling the onset of outbreaks, typically start in small, clustered regions. However, existing literature in this domain is limited by fails to justify rigid distributional assumptions, other covariate effects, and ST data correlation. Addressing these limitations, the Exponentially Weighted Spatial LASSO (EWSL) approach has been developed. This method begins with establishing a baseline model using historical data, such as from 2014-2015, to describe the in-control ST pattern. The current observed data, like that of 2016, is then standardized and decorrelated for real-time monitoring. The EWSL model signals alerts when it detects deviations, demonstrating its effectiveness in accommodating the complexities of ST data, evidenced by small residuals and its success in identifying upward shifts in process mean, which would generally indicate disease outbreaks. This approach has significantly impacted disease surveillance and outbreak detection. Notably, it has reclarified the spread date of H7N9, suggesting an earlier spread than documented in existing literature, and detected early signals that may indicate the emergence of COVID-19. Furthermore, the model observed an increasing trend in influenza from 2014 to 2020.

4 - Joint Modeling of Multistate and Nonparametric Multivariate Longitudinal Data

Lu You, University of South Florida, Tampa, FL, United States, Falastin Salami, Carina Törn, Åke Lernmark, Roy Tamura

It is oftentimes the case in studies of disease progression that subjects can move into one of several disease states of interest. Multistate models are an indispensable tool to analyze such data. In this presentation, a joint model for simultaneous inference of multistate and multivariate nonparametric longitudinal data is proposed to analyze data and answer the research questions brought up in The Environmental Determinants of Diabetes in the Young (TEDDY) study, which is an observational study of at-risk children from birth to the onset of type-1 diabetes. The proposed method allows us to make statistical inferences, test hypotheses, and monitor future disease progression. The performance of the proposed method is evaluated by simulation studies. The proposed method is applied to the motivating example to demonstrate the capabilities of the method.

WB03

Summit - 322

Point Cloud Analytics for Advanced Manufacturing Applications

Invited Session

Quality, Statistics and Reliability

Chair: Yujing Yang, The University of Texas at Arlington, Arlington, TX, United States

Co-Chair: Chen Kan, The University of Texas at Arlington, Arlington, TX, United States

1 - Point Cloud Representation Learning via Graph Signal Processing for Quality Assurance of Additive Manufacturing

Yujing Yang, The University of Texas at Arlington, Arlington, TX, United States

Recent advancements in 3D scanning technologies enable the capture of high-resolution 3D point cloud data, which provides a great opportunity for the detection of geometric defects and surface anomalies in additive manufacturing (AM) products. However, it is challenging to process layer-wisely scanned point clouds and extract defect-pertinent information, due to 1) the layer-to-layer geometric variations and 2) the sparsity of surface defects. To address these challenges, this study develops a novel point cloud representation learning approach based on graph signal processing. By representing point clouds as graphs and the adaptive learning of graph filters, this approach effectively detects both surface and geometric defects in the AM process. The effectiveness of the developed methodology has been demonstrated by both simulation and real-world case studies.

2 - Extracting Features from 3D Point Cloud for Design-Process Knowledge Discovery in Advanced Manufacturing

Zhaohui Geng, Ohio University, Athens, OH, United States

With the advancement of Industrial 4.0, three-dimensional (3D) point clouds have become ubiquitous in many applications, especially those related to the modern manufacturing industries. However, the lower quality of the models generated by 3D scanning or AM constrains the broader adoption of both technologies by the manufacturing industry. Moreover, it is critical to understand the relationship between geometric features, described by the point clouds, and the manufacturability of the advanced manufacturing techniques. The major bottleneck is the lack of suitable analytical methodologies to both fit their unique process characteristics and process the 3D point clouds to extract intricate design-process knowledge from the models produced by both processes. The objective of this presentation is to develop a novel machine learning-driven framework that can extract the process patterns, both global and local, concealed in the point cloud and, combined with the field knowledge, unmask and understand the intricate relationship between geometric design and advanced manufacturing processes. Case studies in additive manufacturing are presented to demonstrate and validate the proposed method.

3 - Pointsgrade: Sparse Learning with Graph Representation for Anomaly Detection by Using Unstructured 3D Point Cloud Data

Juan Du, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, People's Republic of

Surface anomaly detection by using 3D point cloud data has recently received significant attention. To completely measure the common free-form surfaces without loss of details, advanced 3D scanning technologies, such as 3D laser scanners, can be applied and will produce an unstructured point cloud. However, this irregular data structure poses challenges to anomaly detection, in that the existing methods based on regular data, e.g., 2D image, cannot be directly applied. This article proposes a sparse learning framework with a graph representation of the unstructured point cloud for anomaly detection (PointSGRADE). Specifically, the free-form surface is assumed to be smooth. Then, the associated point cloud can be represented as a graph. Subsequently, considering the sparse anomalies, we propose a sparse learning framework and formulate the anomaly
detection problem as a penalized optimization problem, which is further solved by a computationally efficient majorization-minimization framework. Case studies demonstrate the accuracy and robustness of the proposed method. This article proposes a novel methodology for sparse anomaly detection on smooth free-form surfaces represented by unstructured point cloud, which is critical for quality inspection in manufacturing and other application areas.

4 - Filtered Kriging: Improve Kriging Interpolation for Periodic Manufacturing Surfaces with a Pre-filter

Zhiqiao Dong, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Sixian Jia, Chenhui Shao

In modern manufacturing, high-resolution characterization of surfaces is essential in various quality control tasks, including surface quality inspection and tooling maintenance. However, direct high-resolution surface measurements are often prohibitively expensive and/or time-consuming. Interpolation methods are commonly used to obtain denser estimations from more affordable, low-resolution measurements. When 3-D point cloud data is collected over a surface area, it generally corresponds to 2.5-D measurements taken from non-uniformly distributed sites. To interpolate these scattered measurements, methods based on spatial process models, especially kriging-type techniques, are frequently employed. Periodic spatial correlations, which are common in manufacturing applications, cannot be adequately captured by conventional spatial models, thereby causing potential performance degradation or numerical issues. To address these challenges, we introduce a procedure termed filtered kriging (FK), which separates the periodic component using a bandpass pre-filter, allowing the residual to be well fitted with standard models. Drawing on signal processing theories, frequency-domain analyses are conducted to establish effective criteria and devise a practical bandpass filter design strategy. The effectiveness of FK is validated through case studies involving real-world periodic surfaces from two-photon lithography and ultrasonic metal welding. FK has been shown to capture spatial correlation more adequately than conventional methods, and achieves better interpolation accuracy.

WB04

Summit - 323

Data-Driven Urban and Social Analysis

Contributed Session

Chair: Jin Fang, Clark University

1 - Unraveling Seattle's Travel Patterns Through Data Science Exploration

Yangyang Wang, University of Michigan, Ann Arbor, MI, United States, Tayo Fabusuyi, Yifei Sun

This study explores Seattle's travel patterns through a data-driven approach. We analyze social demographic factors, origin-destination pairs, and transportation modes to predict travel behaviors. Employing machine learning models and network analysis techniques, we uncover insights into travel behavior and network structures. Additionally, we develop a real-time user interface for interactive exploration. Validation with real-world data and population synthesis further enhances the robustness of our findings. Our study contributes to understanding travel dynamics and informs policy-making for transportation infrastructure in urban areas.

2 - Enhancing Road Safety: A Data-Driven Spatial Typology of Crashes in New England

Mohammed Abdalazeem, University of Massachusetts Amherst, Amherst, MA, United States, Jimi Oke

This research seeks to address the urgent need for a deeper understanding of roadway crash patterns to improve multimodal safety in New England. We are developing a data-driven spatial typology of crashes at the census tract level, drawing on a comprehensive dataset that integrates crash records, network topology, socioeconomic indicators, and travel behavior metrics from all six New England states. Using dimensionality reduction and clustering techniques, we identify distinct crash characteristic patterns, and reveal geographically-specific variations in crash types. Our analysis uncovers key differentiating factors, such as crash severity, contributing factors (e.g., speeding, weather conditions, roadway design), and modal involvement (e.g., pedestrian, cyclist, passenger vehicle). This research will ultimately produce a predictive model for classifying crash typologies and a publicly accessible database and interactive dashboard. These tools will facilitate data exploration and evidence-based decision-making for transportation safety practitioners, allowing for the development of targeted, data-driven strategies to mitigate crashes and improve safety outcomes.

3 - Analyzing Factors Influencing U.S. City Rankings Over Two Decades: A SEDEA and MRA Approach

Jin Fang, Clark University, Worcester, MA, United States, Yasamin Salmani, Fariborz Partovi

In this study, we analyzed the impact of various social, economic, and environmental factors on the evolution of city rankings over time. We collected data on major U.S. cities over a 20-year period and utilized the super-efficiency data envelopment analysis (SEDEA) model to determine the city rankings for 2000, 2005, 2010, 2015, and 2020. By examining the shifts in these rankings, and conducting a sensitivity analysis through multiple regression analysis (MRA), we identified the most impactful factors, such as cost of living, education, and income. The results significantly inform decisions for individuals choosing where to live and assist policymakers in enhancing urban socio-economic attributes. Additionally, the study provides businesses with essential insights for strategic planning and sustainable growth by identifying key contributors to the city livability index.

WB05

Summit - 324

Optimization and Reliability in Engineering Systems

Contributed Session

Chair: Akshay Mutha, University of Vermont, Burlington, VT, United States

1 - The Price of Complexity in Experimentation

Yudi Huang, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Sebastien Martin, Tony Qin

Tech companies rely on continuous and frequent experimentation to enable rapid system updates. The goal is to implement changes that improve the system, as measured by the experiments' treatment effect. However, these changes may also increase the complexity of the system. For example, experimentation can be harder in the future due to increased potential interferences, or innovation can be slowed down due to technical debt, hurting the company in the long run. The issue is that this effect is hard to model and often overlooked in practice and literature. We use a Markov Decision Process (MDP) to model the development of a tech company as a sequence of experiments and system update decisions. Some changes may be very profitable in the short term but costly in the long term because of increased complexity, which makes the company's decision not trivial. Analyzing optimal policies, we uncover potential traps that lead to significant innovation slowdown, even if firms are aware of complexity issues and even if they can roll back problematic changes. We also show that measuring and anticipating complexity effects, even in a very imperfect way, can significantly help in the long run.

2 - Learning parametric valid inequalities for mixed-integer linear optimization

Shivi Dixit, University of Minnesota Twin Cities, Minneapolis, MN, United States, Qi Zhang

Efficient and safe process operations require decision-making in real-time. Many online decision-making frameworks involve solving discrete optimization problems, many of which present themselves as mixed-integer linear programs (MILPs). However, often the computational complexity of these MILPs presents a major challenge such that the long solution times render them ineffective in time-critical applications. To tackle this challenge, we previously proposed a data-driven surrogate modeling approach, where surrogate models for the MILPs were trained offline in a decision-focused fashion. These surrogate models are linear programs (LPs) and can be solved more efficiently online. While these LP surrogate models prove to be very efficient in providing near-optimal solutions, they are not guaranteed to preserve the true optimal solutions or to even directly predict feasible solutions.

In this work, we build upon our previous work and propose a method for learning effective valid inequalities offline. This is inspired by the cutting-plane method commonly used to solve MILPs, where adding valid inequalities (or cuts) results in tighter MILP formulations. In state-of-the-art MILP solvers, these cuts are added "on the fly" and need to be generated from scratch for every new instance. In contrast, the learned inequalities in our proposed framework are parametric in the input parameters such that they get automatically updated with every new instance. The resulting tighter MILP formulation can then be employed online where it is expected to solve more efficiently while obtaining the same optimal solution as the original MILP.

3 - Surrogate-Assisted Non-linear Economic Model Predictive Control (MPC) for Large-Scale Complex Industrial Semi-Batch Process

You Peng, The Dow Chemical Company, Columbus, OH, United States, Ivan Castillo, Ricardo Rendall

Semi-batch processes have been widely used in industry to produce high-value products. Economic model predictive control (MPC) can be implemented to maximize profits in real-time. However, MPC requires process models which can be difficult to obtain for semi-batch systems due to their non-linear and time-varying dynamics. Often times, we end up with a large system of nonlinear differential-algebraic equation (DAE) that are computationally challenging to solve. Therefore, having simpler but also accurate models can be extremely helpful.

Historically, linear state space models were used as the simplest form of approximation to non-linear process dynamics. However, this requires periodic updates on the controller around the linearized operating points and are only suitable for non-stiff systems. With the emerging of machine-learning based data-driven approaches in the past decade, neural networks (NN) especially structures like LSTM and Transformers that are designed for sequential data modeling, has become the most popular approach for developing dynamic surrogate models.

In this work, we compared a non-linear economic MPC implementation for a large-scale complex semi-batch chemical process at the Dow Chemical Company using 3 types of process models: a DAE based fundamental model with estimated kinetic parameters from process data, a linear state space model fitted from process data and a neural network based dynamic surrogate model trained using both process and simulation data generated from the fundamental model. The result indicates by using an effective surrogate such as NN, similar controller performance when complex DAE model is used can be achieved with much less computational resources.

4 - Product Acquisition for Remanufacturing

Akshay Mutha, University of Vermont, Burlington, VT, United States, Zhaolin Li, Daewon Sun, Jennifer Ryan

We develop models to analyze the acquisition of used products for remanufacturing operations. We perform numerical analyses to show the applicability of our models.

WB06

Summit - 325

Advancements in Edge Computing and Networks

Contributed Session

Chair: Junghyun Kim, Ajou University, Suwon-si, N/A

1 - Deep Network Compression using Kernel Machines

Junghyun Kim, Ajou University, Suwon, Korea, Republic of, Hyunjung Shin

Recently, there has been a surge of interest in On-device AI, which enables Deep Neural Network (DNN) to operate efficiently on small devices (edge devices). For this, it is essential to compress the DNN into a lightweight model while maintaining performance comparable to the original model. One of the leading methods is pruning, which is used to compress models by removing less useful hidden nodes from the network. However, identifying redundant or irrelevant hidden nodes is still ambiguous. In this study, we propose a novel DNN compression method. First, the hidden nodes are transformed to transitional vectors and fed to the kernel machine. Second, a kernel machine plays a role of discerning critical nodes by checking node variability, which measures how sensitive the output values of the node changes depending on the samples. The experiments on benchmark datasets show that the proposed method maintains performance even when the original DNN model is reduced in size by more than 80%.

2 - Deployment Algorithm for End-to-End Time-Constrained Microservices on Geographically Distributed Post-5G Infrastructure

Hirotaka Kasahara, SoftBank Corp., Tokyo, Japan, Ryuichi Kitajima, Jun Towada, Osamu Sato, Takahiro Hirofuchi

This research proposes an efficient deployment algorithm for end-to-end (E2E) time-constrained microservices in the context of edge computing. Edge computing enables real-time data processing by performing computations near the data source, reducing network congestion and latency. The optimal placement of latency-sensitive microservice-based applications poses a challenge in edge computing environments. Such applications, including autonomous driving technology and sensor networks, require strict Quality of Service (QoS) requirements for processing speed and latency. To meet these requirements, it is crucial to minimize data transfer time from IoT devices to edge nodes and ensure responsive communication. To address this challenge, we propose a deployment algorithm that considers the unique characteristics of post-5G infrastructure. This infrastructure includes geographically distributed data centers, Multi-access Edge Computing (MEC), and a highly connected, low-latency, and reliable network. This paper focuses on defining a graph abstraction representation for microservices-based applications with E2E time constraints and formulating the deployment problem as an integer linear optimization problem. We propose a novel deployment technique called “sequential sub-path graph mapping”, which partitions the application graph to accelerate microservices placement calculations. The proposed approach’s effectiveness is demonstrated by comparing it with mixed-integer linear solvers, genetic algorithms, and heuristic methods. The proposed algorithm efficiently produces the application deployment solution that meets resource and E2E time constraints. The proposed algorithm exhibits superior performance in terms of resource utilization costs and minimizing the number of hops between microservices, outperforming existing heuristic methods and metaheuristic techniques.

3 - Decentralized Scheduling in Edge Computing Paradigms: End-to-End Decision Analytics and the Price of Anarchy

Jiayi Zeng, The University of Hong Kong, Hong Kong, China, People's Republic of, Wenjuan Hou, Qiao-Chu He, Zuo-Jun Shen

In this article, we design an edge computing system wherein a platform leverages incentive instruments to coordinate request-heterogeneous computing tasks and resource-heterogeneous edge servers. We propose an integrated optimization model in which the platform decides rewards and scheduling policy, and edge servers individually decide on service capacities, while each computing task minimizes its aggregate cost by selfish routing via available edge servers. We focus on system efficiency evaluation by measuring the Price of Anarchy of the entire system (PoAw), while the efficiency loss due to selfish routing is measured by the Price of Anarchy (PoA). At a strategic level, we decompose the PoAw by analyzing the double marginalization between the platform, edge servers, and computing tasks, which is significantly magnified by PoA. At an operational level, we propose a local-first, local-increasing-costs-and-nonlocal-decreasing-efficiency (LFID) scheduling policy to reduce PoA, with an interesting feature of reversed priorities between local and non-local tasks. In particular, we demonstrate the optimality of the LFID policy for general communication network topologies. The integrated optimization model is solved via reformulations based on SAA. To derive the equilibrium of the selfish routing game, two end-to-end (E2E) computational frameworks are proposed based on neural networks, which produce load data and facilitate solving the integrated optimization model. Case studies are conducted by combining both simulations and real data, and our results advise platform and service providers on scheduling policy design in the edge computing paradigm.

WB10

Summit - 330

Insights into Food Supply Chain and Food Insecurity

Contributed Session

Chair: Suzan Alaswad, Zayed University, Abu Dhabi, United Arab Emirates

1 - Optimizing Store Order Demand Forecasting for Distribution Centers through Data Aggregation Techniques for Store Prepared Items

Sara Branson, Afresh, Mead, CO, United States

Grocery stores rely on distribution centers to supply them with food, and these distribution centers, in turn, source their products directly from farms and food vendors. We use store order forecasts to determine the expected orders that will come into the distribution centers. This enables us to maintain only needed inventory at distribution centers, extending shelf lives and reducing food waste. In the process of forecasting store order demand for distribution centers, we faced difficulties in predicting the demand for products that are ordered in one form and sold in another. While some items are sold both in their original form and in a different form, others are exclusively sold in different forms. For instance, pineapple is sold both as a whole fruit and in sliced pineapple bowls. Another example is meat, where the store orders a sub-primal and sells it in various cuts but very rarely as it is. Current state of the art forecasting techniques fail to consider the contributions from store prepared components. To address this issue, we conducted experiments by using various data aggregation techniques to identify the one that yielded the highest forecast accuracy. We will discuss our approach and the results of our experiments, which improved the forecast accuracy of store order demand for distribution centers.

2 - Developing Data-Driven Market Intelligence Systems for Food Supply Chains

Xaimarie Hernandez Cruz, Arizona State University, Tempe, AZ, United States, J. Rene Villalobos, George Runger

In recent years, the traditional food supply chain has been scrutinized for various reasons, including its high food waste and lack of flexibility in reacting to disruptions in a timely manner. This work discusses an innovative, intelligent machine learning-based system that analyzes the fresh produce market in the United States (US). This system can collect and process relevant fresh produce data, monitor it to identify disruptions, verify the signals, and forecast market prices for any location within the US. Statistical monitoring methods, machine learning pattern recognition techniques, and transformer-based spatial forecasting and interpolation methods were developed to achieve the system's goals. By utilizing this system, supply chain participants will receive timely alerts about market disruptions, enhancing the flexibility of the supply chain and reducing food waste significantly. In this presentation, we discuss the developed market intelligence system and illustrate its components through case studies.

3 - Food distribution improvement via market design approach with optimization-based choice rule

Hou-An Chen, North Carolina State University, Raleigh, NC, United States, Umut Dur, Osman Ozaltin

In this study, we utilize market design tools, particularly matching theory, to offer a market-oriented approach to food distribution through food banks. We examine a market where a single agency can obtain goods from several distribution centers. Our formulation differs from previous studies as we elicit the actual preferences of the agencies for food bundles and incorporate accounts from multiple distribution centers. Based on the renowned Gale Shapley algorithm, our proposed mechanism includes choice rules for distribution centers. In particular, these choice rules are developed using optimization methods that consider three objectives: efficiency, equity, and effectiveness.

4 - When Does Food Become Surplus? A Game Theoretical Model for Partnerships in Food Supply Chains

Suzan Alaswad, Zayed University, Abu Dhabi, United Arab Emirates, Sinan Salman

Our paper presents a game theoretical model designed to examine the dynamics of business partnerships between commercial food supply chains and charitable organizations engaged in food redistribution. The primary objective of this model is to provide decision support and analytical tools for managing food effectively, by maximizing profitability for commercial organizations, minimizing costs for charitable organizations, and facilitating surplus recovery and redistribution within food supply chains. We explore the impact of uncertainties related to food freshness and consumer demand, as well as the effects of cost and operational parameters. Our model's solution, coupled with experimental results, sheds light on the intricate interplay between these factors and the decisions made by the involved parties. Our findings reveal a freshness threshold for the optimal decision-making process as food ages within the supply chain. Furthermore, we demonstrate that food age variance plays a pivotal role in the decision-making process. Higher variance in food age leads to a reduced value of the food for both commercial and charitable organization. These insights highlight the need for improved food age estimation technology and enhanced production, transportation, and storage conditions to minimize food age variance. By quantifying the benefits of such investments, our results provide valuable guidance for practitioners in the food supply chain industry, enabling them to optimize their operations while minimizing food waste. The implications of our research extend to various stakeholders, including government, semi-governmental, and commercial entities.

WB11

Summit - 331

Assortment Planning

Contributed Session

Chair: Sang Won Kim, KAIST College of Business, Seoul, Korea, Republic of

1 - Platform Promotion and Sellers' Price Competition in e-Markets

Sang Won Kim, KAIST College of Business, Seoul, Korea, Republic of, yiying zhang, Youngsok Bang, bosung kim

We consider an e-commerce platform where both the platform and individual sellers provide promotions. We examine how platform-level promotions reshape the buyer structure on the platform, the price competition between individual sellers, and the inequality in sellers' sales. We build an analytical model that considers buyers' heterogeneity in price sensitivity and searching behaviors, as well as sellers' heterogeneity in attractiveness. Our results indicate that as platform promotions decrease the ratio of directly-purchasing consumers to price-comparing consumers on the platform, larger sellers increase their discount amounts faster than smaller sellers. Consequently, platform promotions exacerbate market inequality. We also conduct empirical investigations that provide evidence supporting our analytical findings.

2 - When to Push Ads: Optimal Mobile Ad Campaign Strategy under Markov Customer Dynamics

Guokai Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Pin Gao, Zizhuo Wang

We investigate a seller's optimal advertising campaign strategy targeting customers who interact with the seller over time. We model customers' engagement as a continuous-time Markov chain with two states, active and inactive. While in the active state, customers make purchases according to a Poisson process, with each purchase yielding a specific reward; in contrast, customers in the inactive state make no purchases. The seller can use advertisements to activate customers and the activation probabilities may be affected by customers' fatigue to promotion. The objective of the seller is to maximize the long-term expected average profit by designing an optimal ad campaign strategy based on customers' purchase histories.
For a single customer without budget constraints, we demonstrate the optimality of a triple-threshold policy based on the elapsed time since the last purchase or ad campaign. Moreover, we find that the seller tends to push ads earlier to customers with high purchase rates and low recapture rates (i.e., low transition rates from an inactive state to an active state). Surprisingly, we also find that the seller should push ads earlier to customers with intermediate churn rates (i.e., intermediate transition rates from an active state to an inactive state) compared to those with small or large churn rates.
When confronted with a budget constraint on the overall ad cost, we provide an asymptotically optimal poilcy that first solves a relaxed problem and then implements a separate threshold policy for each customer with a specified budget.

3 - Dynamic Assortment Selection and Pricing with Contextual Learning

Yigit Efe Erginbas, UC Berkeley, Berkeley, CA, United States, Thomas Courtade, Kannan Ramchandran

We consider a dynamic assortment selection and pricing problem in which a seller has N different items available for sale. In each round, the seller observes a d-dimensional contextual preference information vector for the user, and offers to the user an assortment of K items at prices chosen by the seller. The user selects at most one of the products from the offered assortment according to a multinomial logit choice model whose parameters are unknown. The seller observes which, if any, item is chosen at the end of each round, with the goal of maximizing cumulative revenue over a selling horizon of length T. To the best of our knowledge, we are the first to address the problem of dynamic simultaneous contextual assortment selection and pricing. For this problem, we propose an algorithm that learns from user feedback and achieves a revenue regret of order $\wt{\mathcal{O}}(d \sqrt{K T} / L_0)$ where L_0 is the minimum price sensitivity parameter. We also obtain a lower bound of order \smash{$\Omega(d \sqrt{T}/ L_0)$} for the regret achievable by any algorithm, showing that our regret rate is optimal up to logarithmic factors.

4 - Manufacturer-C2M or Platform-C2M: Comparative Analysis of an Innovative E-Commerce Model

Tonghuan Shi, School of Management, Fudan University, Shanghai, China, People's Republic of, Lin Tian, Yifan Xu

In e-commerce, manufacturers frequently launch new products but repeatedly fail due to mismatches with the market. Taking advantage of big data, platforms begin to mine consumer insights and incorporate them into the development process to drive product success, which is known as the customer-to-manufacturer (C2M) model. In this paper, we consider a setting in which a manufacturer distributes products through an online retail platform. We examine two typical C2M modes: one is that the platform sells data to the manufacturer, who then engages in C2M cooperation and develops new products (i.e., Manufacturer-C2M mode); the other is that the platform retains data and takes charge of all decisions related to C2M innovation (i.e., Platform-C2M mode). Our analysis shows that product prices in Platform-C2M mode are generally lower than those in Manufacturer-C2M mode when both firms succeed in their respective innovation. Besides, when the platform demonstrates a higher relative efficiency than the manufacturer, it will invest more to develop C2M products and thus have a higher success rate and vice versa. Furthermore, the preferences of the platform, the manufacturer, and consumers for the two C2M modes are largely determined by the data volume and moderated by the two firms' efficiency in developing products. Unexpectedly, the manufacturer is more likely to prefer the Platform-C2M mode as his efficiency improves relatively. And the platform's entry is not necessarily beneficial to consumers. Additionally, firms and consumers can only align with each other under the Manufacturer-C2M mode, and markets fitting this profile are common in practice.

5 - Dynamic Assortment with Demand Learning under Nonparametric Choice Model

Yifan Li, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Lei Xie, Chaonan Zheng

We consider a dynamic assortment selection problem. During each period, a customer arrives with a deterministic purchase pattern randomly according to a nonparametric choice model. The company offers a subset of n substitutable products with capacity constraint at K, from which the customer purchases according to the specific purchase pattern. All the potential types of purchase pattern are assumed to be known to the company, while the arrival probabilities of each pattern are unknown, and only the purchase information, rather than customers' type, is observable to the company. The company's objective is to optimize total expected revenue over T periods while learning the arrival probabilities by incomplete information. We design a policy that constructs a linear confidential space for the probability vector, and selects the assortment with highest potential revenue within the confidential space, reaching expected regret of order at O(√(κKT logκT)), where κ is the maximum number of how many combinations of arrival probabilities appear in the data set. In some cases mentioned in literature, κ can be proved not to be too large. In order to improve the solvability, we adopt the LP-relaxation on the assortment to solve the assortment optimization problem on [0,1]^n rather than on {0,1}^n and randomize the relaxed assortment back to {0,1}^n. Although it sacrifices the regret performance at O(√nKT+√(κKT logκT)+√nK T), where the O(√nK T) term is the penalty of relaxation, a boundary condition with relaxed assortment falling on {0,1}^n allows to drop the O(√nK T) term and ensures a sublinear regret.

6 - Causal Inference for Pricing of Airline Seats: Evidence from a Large Online Field Experiment

Ravi Kumar, PROS Inc, Houston, TX, United States, Shahin Boluki

We consider the problem of dynamic pricing of a product in the presence of feature-dependent price sensitivity. Developing practical algorithms for robustly estimating price elasticities, especially when information about no purchases (losses) is not available, is a challenge faced by many industries. Based on the Poisson semi-parametric approach, we construct a flexible yet interpretable demand model where the price related part is parametric while the remaining (nuisance) part of the model is non-parametric and can be modeled via machine learning (ML) techniques. We apply the double/debiased ML estimation methodology which makes the estimation of the price-sensitivity parameters robust to biases in the estimators of the nuisance parameters of the model and show the performance of the proposed estimation schemes on simulated and real data from the Airline industry. We also present results from a large scale comprehensive A/B test, which demonstrates that this methodology can substantially improve airline revenues through data-driven price optimization.

WB12

Summit - 332

Assortment Planning with Multiple Warehouses

Contributed Session

Chair: Jayendran Venkateswaran, Indian Institute of Technology Bombay, Mumbai, India

1 - Warehouse Slotting Optimization

Jayendran Venkateswaran, Indian Institute of Technology Bombay, Mumbai, India, Meet Joshi

Efficient warehouse operations are critical in today's consumer goods supply chains, spurred by the tremendous growth in e-commerce. The key warehouse decisions explored in this study include slotting and routing. Slotting refers to the identification of dedicated bins/rack positions for a product or family of products. Order routing refers to the movement of personal/machines within the warehouse to pick-up multiple products of an order. A generic single warehouse consisting of five shelves, each with 25 columns and four levels, is considered. This study explores various algorithms to optimize slotting and routing processes of a warehouse, in response to dynamic customer demands.

The slotting phase employs clustering techniques such as KMeans, HAC, and Pairwise Clustering to identify groups of products frequently ordered together and is applied over single level slotting. For multilevel slotting, the Golden Section slotting strategy is used to predefine slots for stock keeping units (SKUs), and Genetic algorithm for finding optimal slotting.

Effectiveness of slotting strategies is further studied by comparing the total distance travelled and total time required for order picking using various routing options including S-shape routing, Return routing, Mid-point routing, and Travelling Salesman Problem (TSP). Additionally, Simulated Annealing meta-heuristics over TSP is used to optimize travel distance for multiple order picking. Computational experiments, with multiple replications, for slotting of 500 SKUs and with approximately 100 orders for pickups, varying in size from 2 to 50 SKUs per order, have been conducted. Insights into improving warehouse operations through improved slotting and routing are presented.

2 - Inventory Management with Actual Palletized Transportation Costs and Lost Sales

Willow Yang, Sam Houston Sate University, Huntsville, TX, United States, Kanglin Liu, Juan Zhang, Pamela Zelbst

Pallets are widely used for bulk shipping. In response to the call for more comprehensive research on inventory management that incorporates practical freight rates, this study investigates optimal inventory replenishment decisions by explicitly considering palletized shipping rates and lost sales in multi-product multi-period inventory system facing stochastic demand. We first introduce a stepwise transportation cost function that accurately captures real-world palletized less-than-truckload (LTL) and truckload (TL) shipment costs by accounting for both pallet quantity and truck capacity. We then formulate two novel inventory models that consider palletized freight costs and the coexistence of LTL and TL shipments, with one model excluding lost sales and the other specifically incorporating lost sales. We apply both models to a wholesaler in the U.S. and analyze the effects of varying shipping rates and lost sales on inventory decisions. Results show the inclusion of lost sales in the inventory replenishment model can potentially reduce total costs for the company. Interestingly, our findings contradict prevailing industry practices of maximizing truck capacity, revealing that filling up the "free" space on a truck may increase total logistics costs. This holds true even when shipping rates are relatively high and lost sales are considered. Furthermore, our results suggest that joint optimization of transportation and inventory decisions may not be necessary when shipping costs are small relative to inventory costs and lost sales are not considered. However, as shipping rates increase and lost sales are factored in, a joint optimization that accounts for all factors can yield much greater savings.

3 - Multiple-stage solution of the joint order batching and routing problem considering picker blocking for a mixed-shelves warehouse.

Jean-Raymond Fontin, National Taiwan University of Science and Technology, Taipei, Taiwan

Mixed-shelves storage strategy (or scattered storage strategy) has gained a lot of attention in the last 5 years in the literature due to shorter lead times of orders in e-commerce. Shops are forced to considerably decrease the picking time of items to fulfill orders on time. One of the solutions consists of placing Stock-Keeping-Units in multiple locations around the warehouse for greater accessibility. The strategy is called the mixed-shelves storage strategy. Solving the routing problem in a mixed-shelves warehouse is more complex than in other warehouses. When multiple pickers are involved simultaneously the problem becomes even more complex. In this work we present a multiple-stage solution of the joint order batching and routing considering real-life factors such as picker blocking for a mixed-shelves warehouse.

4 - Beyond Traditional Transshipment: A Flexible Mechanism for Multi-Echelon Logistics System

Zheng Tan, University of Science and Technology of China, Hefei, China, People's Republic of, Lindong Liu, Hanzhang Qin, Yugang Yu

Multi-echelon warehousing and logistics systems often incur high costs, particularly for heavy cargo that is inconvenient to store separately. Data from a home appliance logistics company shows that over 54% of orders require transshipment between warehouses, leading to significant operational costs. To address these challenges, we propose a Flexible Transshipment Vehicle (FTV) mechanism. The FTV unloads goods from the mainline fleet at multiple transshipment points, enabling flexible delivery and reducing overall transshipment costs. While static scheduling for FTVs has been studied with pseudo-polynomial algorithms, addressing this in a dynamic scenario remains challenging. We apply a distributionally robust stochastic dual dynamic programming (DR-SDDP) framework, which demonstrates promising out-of-sample performance. Numerical results show that introducing a single FTV can significantly reduce operation costs, with dynamic decisions outperforming greedy strategies without considering long-term effects.

WB13

Summit - 333

Supply Chain Models for Ecommerce

Contributed Session

Chair: Narendra Singh, Nazarbayev University, Astana, Kazakhstan

1 - Learning Cannibalization and Complementarity for Omnichannel Next-Basket Recommendation: A Novel Graph Neural Network Approach

Jiawei Chen, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Zufan Liu, Yue Qian, Yinghui Yang

Omnichannel retailing is a strategy that fully integrates all existing channels to offer customers a seamless experience. A core challenge in omnichannel retailing is recommending the next basket of products to customers in different channels to enhance their experience. To address this, we propose a novel graph neural network approach for omnichannel next-basket recommendation. Our modeling framework effectively captures and infers cannibalization and complementarity relationships between products both within and across channels. We evaluate our approach through comprehensive experiments with real-world data from an omnichannel retailer. Our proposed approach significantly outperforms other alternatives in next-basket recommendation performance for each channel. We also analyze the learned item relationships and provide practical insights.

2 - Buy Online and Return in Store via Partnership

Tao Xu, Drexel University, Philadelphia, PA, United States, Wenjing Shen, Joshua Ignatius

Online retailers are facing more challenges than offline retailers when it comes to dealing with customer returns. We study a BORS (Buy Online and Return in Store) model wherein the online retailers propose an agreement with the local offline retailers to offer offline returns to the customers instead of mailing them back through delivery services. Our goal is to seek what kind of return partnership agreements can benefit customers and both retailers. We find out that for the customers, even though they are offered with better return experience, the prices for the products when implementing BORS partnership will be higher than the prices without such policy. We show that the profits for both online and offline retailers may be higher when they implement the BORS partnership compared to the case when the online retailer provides dropbox to collect returns. BORS partnership policy benefits the online retailer by attracting more customers and when return happens, customers who choose to return in the offline store tend to purchase the substitute products which also benefits the profits of the offline retailers. Finally, we show that the per unit price agreement, in which for every returned product in offline store, the online retailers will pay a certain amount of money to the offline store, is only profitable when there are more high value customers. We provide the management insights to online retailers and help them make the decision on whether to adopt the BORS partnership to the offline retailers.

3 - Omnichannel Product Selection and Shelf Space Planning Optimization

Yajing Chen, Shanghai Jiao Tong University, shanghai, China, People's Republic of

As the retail industry is evolving toward an omnichannel paradigm, we study a product selection and shelf space planning problem faced by omnichannel retailers. We aim to maximize aggregated profits across channels by determining the optimal product offerings for the offline and online channels while optimizing the allocation of limited shelf space to these products in brick-and-mortar stores. Furthermore, the showroom effect of physical stores on online sales is also taken into account. We propose a corresponding model that jointly optimizes these decisions. To solve the nonlinear nonconvex model of practical scale, we first reformulate it into a mixed integer quadratic programming by exploring the structural properties. Next, we use the reformulation linearization technique to further improve the computational speed. Our numerical studies validate the efficiency of the proposed solution approach. In addition, we derive some managerial insights, including that retailers should adopt joint optimization and consider how the involved effects and parameters may influence their profits and decisions.

4 - Store brand interaction on e-tailer's showroom strategy

Yitong LIU, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

Numerous e-tailers have implemented their own store brands in an effort to increase their supply chain competitiveness. And to cope with customers information uncertainty due to purchasing online, many e-tailers choose to open offline showrooms where customers can visit the store to try and inspect products. Our research aims to explore how the store brand interacts with the e-tailer's showroom strategy, including the market coverage investment and contract of sharing with other national brands. We consider a game-theoretic model with an e-tailer, with options to introduce its store brand and showrooms, and a national brand manufacturer (NBM), who sells its product through this e-tailer and could choose to display its product in the e-tailer's showroom with e-tailer's cost-sharing contract. We find that the e-tailer would enlarge the showrooms' market coverage if store brand exists when the operation cost of the showroom is reasonably low and the national brand is relatively less familiar to customers. Additionally, the e-tailer would have additional motivation to open showrooms following the introduction of the store brand. Interestingly, the e-tailer would even have increased profit on customers' disutility of online purchasing by carefully designing the cost-sharing contract. When the national brand is well-known, the e-tailer may refuse the NBM's participation. And when its fame is moderate, the NBM can be better-off from joining in the showroom, leading to a win-win outcome. However the e-tailer will exploit all additional profit of the NBM through manipulating the cost-sharing fraction of showroom if the national brand is rarely acknowledged.

5 - Remanufacturing of Consumer Returns by a Retailer: Implications for Supply Chains

Narendra Singh, Nazarbayev University, Astana, Kazakhstan

I analyze a supply chain where a manufacturer sells new products to consumers through a retailer and the retailer also offers remanufactured products from consumer returns. I examine how strategic interactions among the manufacturer, the retailer, and consumers affect the supply chain profits in the presence of remanufacturing of consumer returns.

WB14

Summit - 334

Online Retail Platforms

Contributed Session

Chair: Bartu Arslan, Eindhoven University of Technology, Eindhoven, N/A, Netherlands

1 - Diversification of Definitions of Latent Attirtion by Customer Segment in Non-Contractual Settings for Multi-Channel, Multi-Brand Firms

Giwon Son, Korea Advanced Institute of Science and Techonology, Seoul, Korea, Republic of, Sungho Park

Although prior studies have addressed definitions of customers' latent attrition in non-contractual settings through customer base analysis, there are few discussions on how it can be defined at firm-wide level in multi-channel, multi-brand firms. Since their customers enjoy unique experiences that each brand and channel offers, along with firm-wide benefits, these customers may exhibit highly diverse purchasing and exploring behaviors. Therefore, applying the same fixed time-window across all customers, on which attrition is defined as non-purchases in existing studies, may be inappropriate for our setting. Survival analysis may be an alternative approach, but it could require considerable computational time and costs if the number of observations and/or variables is large. In this study, leveraging transaction, exploration, and experiential data from a beauty and personal care firm, we propose a methodology to define latent attrition, aiming to optimize profitability of the anti-churn marketing campaign by examining customer segments derived from various methods, including the RFM criterion and a machine learning clustering algorithm such as BIFGS. We also provide key factors crucial for predicting attrition in our setting.

2 - Competitive Persuasion with Biased Disclosure

Yue Feng, Tianjin University, Tianjin, China, People's Republic of

Companies often induce consumers to buy by disclosing product information, such as advertisements, audits, and survey reports. Despite the importance of information disclosure, it is crucial to recognize that firms' self-disclosed information is often tainted with bias. Firms tend to exhibit positive bias when presenting information about their own products, aiming to downplay any negative aspects and emphasize the positive attributes. On the contrary, when disclosing information about competitors' products, firms may adopt a negative bias, exaggerating the flaws and concealing the strengths of rival products.

In this paper, we investigate a competitive-information-design variant where two rival firms employ biased information disclosure in a sequential Bayesian persuasion game to influence consumer purchase decisions towards their respective products. The primary contributions of this article are twofold. Firstly, it delves into the endogenous decision-making process of the competing firms, specifically focusing on the critical determination of whether to disclose their own information or that of their competitor. Secondly, the paper presents tailored information design strategies for competitive scenarios, particularly addressing situations where consumers' prior beliefs about the two products differ.

3 - Exploring the Current State and Future Prospects of Servitization in Small and Medium-sized Enterprises: A Follow-up Study

Chulok Ahn, Kunsan National University, Gunsan-si, Jeollabuk-do, Korea, Republic of, Sungyong Choi

Servitization has emerged as a crucial strategy for manufacturers, particularly for gaining a competitive edge in the market. While numerous success stories have been reported, most of the existing studies focus on large enterprises, leaving a gap in understanding its applicability to Small and Medium-sized Enterprises (SMEs). In this follow-up study, we delve into the current state and future prospects of servitization specifically within SMEs. Our focus industries are electrical equipment manufacturing and basic metal parts production, chosen due to their relevance across manufacturing sectors.

Employing a qualitative research approach, we conducted in-depth interviews with CEOs, managers, and employees from selected companies in these industries. The analysis aimed to assess the current level of servitization and identify potential opportunities for further integration of services into their business models. Our findings reveal that the level of servitization in these industries remains relatively low among SMEs, with limited prospects for significant changes. Challenges such as capital constraints and limited human resources hinder the adoption of servitization practices.

However, despite these challenges, companies acknowledge the importance of servitization for long-term sustainability and express willingness to embrace opportunities for growth in the future. By revisiting these companies and observing any evolving trends or shifts in attitudes towards servitization, our study seeks to contribute to a deeper understanding of servitization dynamics within SMEs and provide insights for enhancing their competitiveness in the evolving market landscape.

4 - Optimizing Reusable Packaging Production Decisions and Return Logistics in E-commerce

Bartu Arslan, Eindhoven University of Technology, Eindhoven, Netherlands, Albert Schrotenboer, Zumbul Atan

Increasing environmental concerns and the growth of e-commerce necessitate sustainable practices in the retail industry. In collaboration with an industry partner, we introduce an optimal control framework for managing reusable packaging logistics in online retail. From the perspective of a reusable package producer, we consider a dynamic system integrating multiple packaging sizes, production decisions, online retailers, and delivery services. We focus on balancing the number of reusable packages within the system with the cost of intervening in the return process. The flow starts with the initial production of reusable packages, followed by allocation to retailers, and then delivery to customers. Subcontracted delivery services facilitate the delivery and retrieval of packages, forming a semi-open supply chain network due to potential losses during transportation. The core of our study is the collection decision of packages from delivery depots to the producer's warehouse, incurring large fixed costs for truck dispatches and variable costs for each visited depot. This resembles a stochastic Joint Replenishment Problem (JRP). We model this as a Markov Decision Process, considering sequential production, fulfillment, and replenishment decisions to minimize overall costs, including production, logistics, and inventory costs. To solve this complex problem, we employ Deep Reinforcement Learning and heuristic methods to manage the system's high-dimensional state space. Our solutions aim to outperform traditional approaches, providing near-optimal strategies for reusable package management and offering insights into sustainable practices in online retail logistics.

WB15

Summit - 335

Innovative Applications in Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Rahul Roy, Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Co-Chair: Nur Sunar, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States

Co-Chair: Jayashankar Swaminathan, University of North Carolina Chapel Hill, Chapel Hill, NC, United States

1 - Online Facility Location: Running Stores on Wheels with Spatial Demand Learning

Yan Zhang, McGill University, Montreal, QC, Canada, Junyu Cao, Wei Qi

A shift toward shopping at (autonomous) wheeled vending stores is redefining urban retail. Compared with traditional brick-and-mortar stores, such mobile stores are cost-efficient to deploy and adaptive to fast-evolving business environments. However, mobile stores are confronted with unknown demand and limited capacity. Store mobility enables demand learning and profit maximization, yet an optimal dynamic store location policy remains unclear. We model this "learning-and-earning" problem by taking optimistic actions under parameter uncertainty. The joint optimization over parameter and action set is complicated by the combinatorial nature and infinite choices within the action set. We overcome these challenges by leveraging continuous approximation methods, and then propose a continuous-approximation optimistic (CA-O) learning framework under some special problem structures. Nevertheless, for more general scenarios, the problem remains intricate due to the nonconvexity in unknown parameters. We alternatively propose a CA-O faster learning algorithm by utilizing first-order approximation techniques and further proving a closed-form gradient to guarantee computational efficiency. We theoretically analyze and numerically validate the regret for the proposed algorithms. In a Toronto case study, our algorithm significantly outperforms baselines. Mobile stores earn higher profits than brick-and-mortar stores through demand learning and store mobility. More broadly, this paper envisions the future landscape of urban retail enhanced by omnipresent mobile facilities.

2 - Bulk vs. Sequential Product Returns: Algorithms and Insights for Assortment Planning

Sahika Sahan Konur, Texas State University, McCoy College of Business, San Marcos, TX, San Marcos, TX, United States, Jacob Feldman

The convenience of online shopping has transformed it into a retail powerhouse. However, this shift has significantly increased the frequency of product returns, imposing additional costs on retailers, including refurbishing and restocking. Our paper tackles these challenges by comparing the Sequential Returns model by Wagner and Martínez-de Albéniz (2020) with our newly proposed Bulk Returns model. The Sequential Returns model assumes customers purchase one item at a time, while the Bulk Returns model caters to consumers who order multiple products simultaneously. Each model integrates a returns disutility parameter to quantify the inconvenience of returning items.

In this paper, we focus on the assortment optimization problem under both returns models. The goal is to select a subset of products to offer to maximize expected profits. We show that even without return costs, the assortment problem under the Bulk Returns model is NP-Hard. Moreover, we show that there exists an optimal polynomial-time algorithm for the Sequential Returns model when return costs are homogeneous. Our theoretical analysis indicates the Sequential Returns model generally yields higher expected profits and market share. Numerical experiments using real-world data measure the profit improvement afforded by the Sequential Returns model. We also analyze how this profit improvement compensates for the potentially increased number of return trips in Sequential Returns model, which is capped at one under the Bulk Returns model.

This study provides key insights for optimizing product recommendations and managing return behaviors, offering strategies for retailers to enhance profitability and customer satisfaction amidst return costs and inconveniences.

3 - Impact of Market Thickness and Heterogeneity on Matches in Online Marketplaces with Active Providers

Jayashankar Swaminathan, University of North Carolina Chapel Hill, Chapel Hill, NC, United States, Nur Sunar, Rahul Roy

In online service marketplaces, supply-side thickness (the number of providers) is widely believed to be crucial for facilitating matches, i.e., transactions between providers and customers. The empirical literature generally supports this view, providing evidence for the hypothesis that market thickness increases matches, albeit at varying rates. This support is typically obtained in contexts with a passive seller listing where all sellers are readily listed for customers. Distinctively, our study empirically examines an online marketplace where providers are active, meaning they must extend a proposal to be listed. For our study, we collaborated with a leading U.S. online solar marketplace that connects solar panel installers with potential adopters. We use an instrumental variable approach and advanced clustering algorithm to investigate how supply-side thickness affects transactions in such marketplaces. Our analysis of an online marketplace with active providers reveals results that challenge the conventional understanding: We find a significant, inverted U-shaped relationship between market transactions and supply-side thickness. This indicates that transactions initially increase with thickness up to a certain thickness threshold, after which they begin to decline. This finding diverges from the general understanding in the literature, suggesting the need for new marketplace design strategies. We show that increased market heterogeneity negatively impacts transactions, and this effect is exacerbated in thicker markets. To our knowledge, this is the first empirical study that explores the effect of market heterogeneity on marketplace transactions with active providers. Our findings highlight the complex effects of increased thickness and heterogeneity, offering valuable managerial insights.

4 - Dynamic Decision Making with Operational Data Analytics (ODA)

J Shanthikumar, Purdue University, WEST LAFAYETTE, IN, United States, Qi Feng, Yang Yang

We will discuss the application of Operational Data Analytics (ODA) to Dynamic Decision Making in Revenue Management, Inventory Control and Stopping Time problems.

WB16

Summit - 336

Revenue Management: Sustainable and Fair Approaches

Invited Session

Revenue Management and Pricing

Chair: Mina Iravani Mohammadabadi, Hong Kong, Hong Kong

Co-Chair: Ozge Sahin, Johns Hopkins Carey Business School, Baltimore, MD, United States

1 - Fairness-Aware Assortment and Pricing Optimization under Inventory Constraints

Yanzhi Li, City University of Hong Kong, Hong Kong, China, People's Republic of, Hongbin Zhang, Yu Yang

We study a retail platform that sells products from multiple suppliers with inventory constraints. The platform aims to maximize the minimum expected revenue of all suppliers by personalizing assortments and prices to customers, whose purchase behavior is unknown. Assuming customer choices follow a sequential multinomial logit (SMNL) model, we develop online learning algorithms for various scenarios, considering different planning horizons and inventory bounds. Our algorithms achieve sub-linear regrets in all cases, ensuring fair treatment of brands and improved engagement in the e-commerce environment.

2 - Transparency and Power Dynamics: A Game Theoretic Analysis of the Supply Chain Due Diligence Regulations

Ali Shantia, HEC Paris, Paris, France, Sara Rezaee Vessal, Sam Aflaki

We explore the impact of supply chain sustainability policies, focusing particularly on the roles of asymmetric bargaining power and transparency. Utilizing a game-theoretic model within a two-tier supply chain framework, the study uncovers that the asymmetric power distribution between buyers and suppliers crucially influences sustainability compliance levels. Key findings suggest that a strategic balance between high sustainability standards and corresponding non-compliance penalties is essential for enhancing policy effectiveness. However, the effectiveness of these standards and penalties is contingent upon the relative bargaining power of supply chain actors and the degree of transparency within the chain. Limited transparency typically results in only partial compliance with sustainability policies. Consequently, the optimal formulation of these standards and penalties must be tailored to the specific power dynamics and transparency levels present in the supply chain. For managers, the study emphasizes the necessity of comprehending supply chain dynamics to foster effective sustainability compliance. Moreover, policymakers are advised to consider these dynamics when designing sustainability policies, thereby potentially increasing the sustainability and ethical responsibility of supply chain operations.

3 - Default or Dare: The Impact of Defaults on Consumer Choice, Assortment, Pricing, and Estimation

Jing Peng, University of Science and Technology of China, Hefei, China, People's Republic of, Ming Hu, Ruxian Wang

The default effect, the tendency of individuals naturally to pay more attention to the default option than alternatives. We incorporate the default effect into the classic multinomial logit choice model and study the associated problems. We characterize the default effect by multiplying an enhancement factor with the attractiveness of a selected product. Interestingly, setting a default option is unnecessary to benefit the total expected profit, while optimally setting the default option guarantees an improvement of the total expected profit. We find that the optimal selection of the default option depends on the product markup and attractiveness, as well as the magnitude of the default effect. Moreover, jointly optimizing the default option and assortment shrinks the size of the optimal assortment while jointly optimizing the default option and prices improves the markups of all products. In the extensions, we first consider selecting multiple default options with different enhancement factors and characterize the optimal selection of the default options to maximize the total expected profit. Second, we identify the conditions under which the optimal sequence of default selections is the same for both the firm and consumers. Third, we incorporate product fairness into the default optimization problem and develop an algorithm to find the optimal default option efficiently. Fourth, we characterize the default effect by additive enhancement and show the optimal default option and pricing policy. Finally, we develop a maximum likelihood estimation method and use synthetic data to validate the proposed model and illustrate how to mitigate estimation bias.

4 - Size Substitution and Fairness in Assortment Optimization

Alireza Alavi, McGill University, Montreal, QC, Canada, Mehmet Gumus, Javad Nasiry

In retail settings, customers facing size stockouts may opt to substitute with neighboring sizes. Stockouts also raise fairness concerns, especially among marginalized size groups. We develop static assortment optimization algorithms to find profitable yet fair assortments in the presence of size substitution and fairness concerns.

5 - The Impact of Multi-unit Promotions on Sustainability and Fairness: Waste and Fair Allocation

Mina M Iravani , Hong Kong University of Science and Technology (HKUST), Hong Kong, Hong Kong, Ozge Sahin, Ying-Ju Chen

Sustainability has emerged as a paramount concern in today's world, as urgent environmental challenges like climate change, deforestation, and ocean pollution loom large. Excessive consumption and the resulting waste generation are major contributors to these issues. Unfortunately, common marketing strategies often employed by sellers to incentivize consumers to purchase more mostly exacerbate this problem. While these strategies may enhance equity by increasing consumer access to products, they also pose a threat to the environment. Consequently, striking a balance between sustainability and consumer access becomes a crucial and challenging task. We focus on pricing bundles and analyze its impact on sustainability and equity. We propose a novel strategy that not only decreases the waste and increases consumers' access but also improves the seller's profit. We compare the proposed strategy with take back legislation, analyzing their impact on waste, seller profit, and product accessibility, and outlining the specific conditions under which each approach can be most effective.

WB17

Summit - 337

Advances in (Network) Revenue Management

Invited Session

Revenue Management and Pricing

Chair: Rui Zhang, University of Colorado Boulder, Boulder, CO, 80309, United States

1 - Probabilistic Approximations for Network Revenue Management

Rui Zhang, University of Colorado Boulder, Boulder, CO, United States

We propose probabilistic approximations to capture the interactions among resources for network revenue management. The probabilistic approximations are stronger than SPL approximation in the sense that they lead to tighter upper bounds. Our numerical results also suggest that the probabilistic approximations lead to better control policies than the SPL approximation.

2 - Gradient Descent Method in Network Revenue Management: Improved Results

Sichen Guo, Shanghai University of Finance and Ecnomic, Shanghai, China, People's Republic of, Cong Shi

We address a standard online network revenue management problem, where a firm aims to maximize revenue over a finite selling season without knowing the precise relationship between price and demand rate. A major challenge is the high dimensionality of pricing decisions and the interdependencies between products. To the best of our knowledge, this problem has not been fully resolved. We assume that the revenue function is strongly concave in the price-controlled demand rate, while the decision maker can only adjust prices. Leveraging the idea of online inverse batch gradient descent (IGD) algorithm, we overcome several significant challenges stemming from inverse estimation and the lack of first-order information. Our algorithm and analysis framework can be further extended to limited inventory cases and concave problems.

3 - Network Revenue Management Problem: A Computational Method

Hongzhang Shao, University of Chicago Booth School of Business, Chicago, IL, United States, Baris Ata

We consider a high-dimensional network revenue management problem and develop a simulation-based computational method that relies heavily on deep neural network technology. The effectiveness of our policy is demonstrated through problem instances derived from the airline revenue management dataset provided by Topaloglu (2009).

WB18

Summit - 338

Optimization of Electric Power Systems Under Decision-Dependent Uncertainty

Invited Session

OPT: Optimization Under Uncertainty

Chair: Alexandre Moreira, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

1 - Dynamic Transmission Line Switching Amidst Decision-Dependent Wildfire-Prone Conditions

Juan-Alberto Estrada-Garcia, University of Michigan, Ann Arbor, MI, United States, Ruiwei Jiang, Alexandre Moreira

Power grids are vulnerable to wildfire and its propagation over time. Wildfire-related transmission line failures can trigger extended power in regions of the system. During dry seasons, environmental conditions increase the risk of wildfires and their propagation, exposing power grids to cascading failure conditions. Simultaneously, transmission line operations beyond normal limits, as a response to compensate for failed lines, can cause failures and subsequent fire ignition. In this work, we propose a multi-stage optimization model to dynamically switch transmission lines to adapt the network structure to respond to the wildfire while considering the trade-off between operational performance and wildfire propagation. The proposed formulation is a multi-stage model with decision-dependent probabilities of line failure and fire propagation. We developed a stochastic nested decomposition algorithm to solve this model and present a case study using a 120-bus transmission system.

2 - Leveraging Mdps and Decision-Dependent Uncertainty Sets for Effective Distribution System Operation in Wildfire-Prone Areas

Prasanna Raut, University of Washington, Seattle, WA, United States

Wildfires pose a significant threat to electricity grids, often resulting in prolonged power outages. With climate change increasing the frequency of dry conditions, this threat is expected to escalate. It's crucial to prevent human-induced wildfires, including those triggered by power system operations, especially in areas prone to wildfires. In this work, we propose a Markov Decision Process (MDP) model to capture the relationship between the availability of lines and the line-flow decisions made by a System operator. As the transition probabilities are only known through sample estimates, we need a method that takes into account the ambiguity of these probability distributions. To this end, we employ Distributionally Robust Bellman's equations for obtaining optimal policies for operation of the distribution system. This method ensures that the chosen policy remains robust across various probability distributions within a decision-dependent ambiguity set, thus enhancing the system's resilience to uncertainties in wildfire-prone environments.

3 - Mitigating the Risk of Wildfire Ignitions from Electric Grid Infrastructure

Line Roald, University of Wisconsin - Madison, Madison, WI, United States

The electric grid is known to cause the ignition of highly destructive and deadly wildfires. Our research has considered different methods for managing this risk, most notably de-energization or undergrounding. We will discuss these and other mitigation measures in the context of optimization under decision-dependent uncertainty, as well as challenges (and solutions) for modeling and solving the resulting optimization problems.

4 - Decision-Dependent Uncertainty-Aware Distribution System Planning Under Wildfire Risk

Alexandre Moreira, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

The interaction between power systems and wildfires can be dangerous and costly. Damaged structures, load shedding, and high operational costs are potential consequences when the grid is not properly prepared. In fact, the operation of distribution grids can be liable for the outbreak of wildfires when extreme weather conditions arise. Within this context, investment planning decisions should consider the impact of operational actions on the uncertainty related to wildfires that can directly affect line failure likelihood. By not properly recognizing this aspect, the evaluation of the trade-off between cost and benefit when planning system investments to deal with wildfire risk can be compromised. In this paper, we propose a decision-dependent uncertainty (DDU) aware methodology that provides the optimal portfolio of investments for distribution systems while considering that high power-flow levels through line segments in high-threat areas can ignite wildfires and therefore increase the probability of line failures. Thus, our methodology identifies the best combination of system upgrades (considering the installation of new lines, hardening existing lines, and placement of switching devices) to provide the necessary leeway to operate the distribution system under wildfire-prone conditions. Our case study demonstrates that the load shedding and associated operational costs can be substantially decreased by adequately modeling the DDU relationship between power flow prescriptions and line failures in locations where wildfires can take place.

5 - Assessing Traffic Patterns in Egress Modeling: Identifying Challenges in Evacuation Route Accessibility during Wildfire Events

Fatemeh Aarabi, San Diego Gas and Electric, San Diego, CA, United States

Given the increasing frequency and intensity of wildfires, the development of an effective evacuation model is paramount for ensuring public safety and reducing casualties. This presentation introduces a wildfire evacuation model that integrates fire spread shape with traffic patterns and human behavior during evacuation scenarios. Key parameters, such as wind direction and gusts, are employed to determine the progression of the fire, while the characteristics of the underlying road network facilitate the prediction of evacuation behavior. Additional factors, including demographic variables such as age, are incorporated to refine travel time estimations across the network. The model ultimately identifies critical bottlenecks and high-risk locations, providing projections of the affected population. A real-world case study of San Diego County will be utilized to demonstrate the model's applicability and effectiveness.

WB19

Summit - 339

Energy Systems Planning

Contributed Session

Chair: Kristen Schell, Carleton University, Ottawa, ON, Canada

1 - A sample robust optimal bidding model for a virtual power plant

Seokwoo Kim, Pohang University of Science and Technology (POSTECH), Pohang, Korea, Republic of, Dong Gu Choi

This study explores one operational challenge for a virtual power plant—the optimal bidding for highly uncertain distributed energy resources in a day-ahead electricity market. The optimal bidding problem can be formulated as a scenario-based multi-stage stochastic program. However, the scenario-tree approach raises two consequent issues—scenario overfitting and massive computation cost. This study addresses the issues by deploying a sample robust optimization approach with linear decision rules. A tractable robust counterpart is derived from the incentive-based market structure, a unique feature of the South Korean market. The original model is then reduced to a two-stage stochastic mixed-integer linear program and then efficiently solved by adopting a dual decomposition method combined with heuristics. Computational studies were conducted using real-world business data with several benchmark models.

2 - Convergence Guarantees of a Distributed Network Equivalence Algorithm for Distribution-OPF

Yunqi Luo, University of Science and Technology of China, Hefei, China, People's Republic of, Rebayet Sadnan, Bala Krishnamoorthy, Anamika Dubey

Optimization-based approaches have been proposed to handle the integration of distributed energy resources into the electric power distribution system. The added computational complexities of the resulting optimal power flow (OPF) problem have been managed by approximated or relaxed models, which may lead to infeasible or inaccurate solutions. Other approaches based on decomposition-based methods require several message-passing rounds for relatively small systems, causing significant delays in decision-making. We propose a provably convergent distributed algorithm called ENDiCo-OPF for nonlinear OPF. Our method is based on a previously developed decomposition-based optimization method that employs network equivalence. We derive a sufficient condition under which ENDiCo-OPF is guaranteed to converge for a single iteration step on a local subsystem. We then derive conditions that guarantee convergence of a local subsystem over time. Finally, we derive conditions under a suitable assumption that when satisfied in a time sequential manner guarantee global convergence of a line network in a sequential manner. We also present simulations using the IEEE-123 bus test system to demonstrate the algorithm's effectiveness and provide additional insights into theoretical results.

3 - Optimization of Integrated Energy System Operation under the Virtual Power Plant V2G Model

Yuping Huang, Guangzhou Institute of Energy Conversion, Guangzhou, China, People's Republic of, Tong Ye

The study explores the optimization of integrated energy system operations in industrial parks under the VPP model. A multi-objective optimization model is built to enhance the coordinated operation of power users, load aggregators, power storage units, PV power generation, and electric vehicle charging-discharging and swapping systems. Based on a two-stage scheduling framework with a dual settlement mechanism, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is employed to solve this complex optimization problem. This method provides IES operators with an effective decision-making tool to enhance energy efficiency and reduce operational costs, thereby addressing the volatility and uncertainties of the electric power market.

4 - Integrating thermal and electrical dependencies into optimal community energy planning

Kristen Schell, Carleton University, Ottawa, ON, Canada, Laura Catalina Jaramillo Villarreal

Most energy system optimization models focus only on electrical supply and demand. However, many northern communities have thermal energy needs that are twice those of the electrical side of the energy system. Research indicates that addressing thermal energy solutions is crucial for the sustainability and energy independence of cold climate communities. By prioritizing heating while maintaining electrical supply, energy systems can be adapted to ensure a more balanced and effective energy strategy for these regions. This study develops an energy system optimization model that combines both the electrical and thermal energy demands, as well as their points of interdependency. The optimization model is applied to a case study of an Arctic Canadian community, Aklavik in the Northwest Territories. Results show how incorporating thermal and electrical interdependencies affect optimal investment decisions and community energy plans. In discussing these challenges, recommendations are made for how the model can inform energy policy and decision-making in community energy planning.

WB20

Summit - 340

Incentive Design for Multi-agent Agent Systems

Invited Session

Decision Analysis Society

Chair: Sukanya Kudva, UC Berkeley, Berkeley, CA, United States

Co-Chair: Anil Aswani, UC Berkeley, Berkeley, CA, United States

1 - Optimal Interventions in Coupled-Activity Network Games: Application to Sustainable Forestry

Rohit Parasnis, Massachusetts Institute of Technology, Cambridge, MA, United States, Saurabh Amin

We consider the problem of promoting sustainability in production forests wherein a given number of strategic entities are authorized to manage concession regions. These entities harvest commodities and sell them in a market. Concession owners (agents) engage in producing a commodity that does not interfere with protected forest resources (sustainable activity) but are also inclined to infringe into protected regions to expand their agricultural footprint (unsustainable activity). Current sustainability certification programs aim to incentivize agents to reduce their individual levels of unsustainable activity by providing them pricing incentives to compensate for sustainability certification costs. However, such incentives ignore the coupling between the sustainable and unsustainable activities and the impact of interactions between agents that share concession boundaries and/or production resources. Importantly, they do not guarantee any aggregate reduction in the unsustainable activity. To design optimal pricing policies that address these limitations, we adopt a coupled-activity network game model and characterize two types of policies in a budget-constrained setting: one that maximally suppresses the aggregate unsustainable activity and another that maximizes social welfare while constraining the aggregate unsustainable activity to remain below a predefined tolerance. Our analysis provides new insights on the agents' equilibrium effort levels resulting from intra-activity and cross-activity network effects. We also develop an algorithm that solves our non-convex welfare maximization problem for an important class of pricing policies and real-world concession networks. Besides, we identify a measure of node centrality that helps us determine pricing incentives that can lead to effective suppression of the aggregate unsustainable activity.

2 - Market Mechanisms for Advanced Air Mobility

Victoria Tuck, UC Berkeley, Berkeley, CA, United States

Advanced Air Mobility (AAM) is poised to be a multibillion dollar industry and has the ability to transform mobility by providing new means for urban transportation, cargo delivery, and more. Due to the expected scale and heterogeneity of such operations, the Federal Aviation Administration (FAA) is advocating for market-based mechanisms for strategic management of airspace, to be facilitated by Providers of Services for UAM (PSUs). Leveraging insights from economic theory, we will discuss two mechanisms to be deployed by PSUs which encourage fair and efficient operations while safely allocating limited resources in infrastructure (eg. vertiports). First, we propose an incentive-compatible and individually rational vertiport reservation mechanism that uses the monetary bids from operators to maximize the overall valuations of all operators while improving the congestion at vertiports. Additionally, we improve the computational tractability of designing the reservation mechanism by proposing a mixed binary linear programming approach that is based on constructing a network flow graph corresponding to the underlying problem. Second, we propose an iterative market mechanism, inspired from recent advances in Fisher markets that is based on revealed preferences of operators in the form of congestion on the infrastructure and does not require them to reveal bidding information for their desired paths. Finally, we discuss open challenges and opportunities.

3 - Dynamic Contract Design with Learning

Yong Liang, Tsinghua University, Beijing, China, People's Republic of, Shiyuan Wang, Peng Sun

We investigate a dynamic moral hazard problem in which the agent's capability is unknown to both the principal and the agent in the beginning and can be learned over time. Specifically, the agent can exert effort in order to generate random arrivals that are beneficial to the principal. The probability of arrival, however, is determined by the agent's capability. Not knowing the agent's type as well as whether the agent exerts effort or not makes it extremely hard to identify the optimal dynamic contract. Therefore, we focus on designing dynamic contracts that achieve the best regret rate. In a discrete-time setting with two possible agent's capability types, we propose two types of contracts, both of which achieve the regret rate of $O(\ln T)$ for a time horizon T. Our contracts are ``history-independent,'' and therefore easy to implement for the principal. They motivate the agent to always exert effort before termination, and therefore are easy for the agent to respond to. We also establish a regret lower bound, which is $\Omega (\ln T)$ among all contracts, including the ones that do allow the agent to shirk. This implies that our regret rate is the best possible. We further extend these results to continuous-time settings.

4 - Trilevel Optimization for Hierarchical Decision Making

Sukanya Kudva, UC Berkeley, Berkeley, CA, United States, Anil Aswani

Multi-level optimization problems can help the industry and the government in making decisions that involve a hierarchy of players or a chronology of events in time. The first player in a hierarchy or the first chronological event takes the first action, followed by the next in hierarchy/ chronology, and so on. And at each level, the action optimizes an objective subject to the actions previously taken and is dependent on future actions that could be taken. In this work, we propose an algorithm for a trilevel optimization problem, when there are three such levels. We discuss our model assumptions, algorithm, convergence proof and experimental results.

5 - Designing Service Menus for Bipartite Queueing Systems with Strategic Customers

Lisa Hillas, University of Auckland, Auckland, New Zealand, Rene Caldentey, Varun Gupta

We consider a multiclass, multiserver queueing system, in which customers of different types have heterogeneous preferences over the many servers available. The goal of the service provider is to design a menu of service classes that balances two competing objectives: (1) maximize customers’ average matching reward and (2) minimize customers’ average waiting time. A service class corresponds to a single queue served by a subset of servers under a first come, first served–assign longest idle server service discipline. Customers act as rational self-interested utility-maximizing agents when choosing which service class to join. In particular, they join the class that maximizes their expected ex ante net utility, which is given by the difference between the server-dependent service reward they receive and a disutility based on the mean steady-state waiting time of the service class they join. We study the problem under (conventional) heavy-traffic conditions: that is, in the limit as the traffic intensity of the system approaches one from below. We prove that there are simple service menus that will optimise each of the two service provider objectives separately, and provide three mixed-integer linear programmes that identify a range of service menus that will trade off-the two objectives.

WB21

Summit - 341

Advancements in Healthcare and Service Analytics

Contributed Session

Chair: George Miller, Altarum Institute, 3100 E Delhi, Ann Arbor, MI 48103, Ann Arbor, MI, 48103, United States

1 - Usage Analysis of the Southeast Wisconsin Paratransit System

WILKISTAR OTIENO, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

To protect the rights of seniors and Persons with Disabilities (PWDs), the United States government, through Congress, enacted the Americans with Disabilities Act (ADA) in 1990. The Act ensures that there is equity for persons with disabilities in all aspects of public services including employment, education, transportation, accommodation, commercial facilities and businesses and communication (Federal Register, 2010). In this study, we focus on the problem of transportation of PWDs, specifically in Milwaukee County Transit System (MCTS). The research first explores the effect that increased ADA bus ridership would have on the paratransit system performance such as bus utilization by PWDs, PWD average waiting time as well as the average time in system. Per the results, the average utilization was found to be 6.5%. This percentage is an indicator of that MCTS has potential to increase fixed bus route ridership by persons with disabilities, especially if challenging issues such as low bus frequency, less geographical coverage of the bus network (to cover areas where most Casino ADA visitors reside), public transport awareness, bus driver training and most of all, increased accessibility at the destination.

2 - Bayesian Analysis of Hospital Service Quality:Pre- and Post-COVID

Ling Hu, University of South Florida, TAMPA, FL, United States, Sajeev Varki, Daniel Zantedeschi

The hospital sector is the largest revenue-generating industry in the United States. Its service quality profoundly influences societal welfare. The Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), administrated by the CMS (Centers for Medicare &

Medicaid Services), serves as a cornerstone in fostering patient-centered care, enhancing informed decision-making, and optimizing healthcare resource allocation.

Departing from the conventional methodologies using linear or multilevel linear regression models with Maximum Likelihood Estimation (MLE) to analyze cross-sectional rating score data, this research employs a novel approach. Leveraging a multilevel Item Response Theory (MLIRT) model, it delves into the hierarchical structure of HCAHPS data through Bayesian estimation. By separately estimating the latent hospital service quality and item parameters of threshold and discrimination, the MLIRT model effectively addresses measurement error, enabling more precise estimation. Bayesian estimation is favored over MLE due to its superior ability to incorporate prior knowledge or beliefs about these parameters and handle heterogeneity issues.

Our study provides insights for hospitals to optimize resource allocation to improve service quality. Our findings offer actionable recommendations by identifying 'Care Transition,' 'Nurse Communication,' and 'Doctor Communication' as the foremost service attributes reflecting latent service quality. Furthermore, our investigation into the service quality change before and after COVID-19 provides hospitals and policymakers with a more comprehensive understanding of the pandemic's impact, which may help develop appropriate strategies when facing similar crises. Lastly, our research introduces an alternative framework for evaluating hospital service quality, offering policymakers a fresh perspective on assessment methodologies.

3 - Nursing Home Staffing Shortages and the New Federal Nursing Home Standards

George Miller, Altarum Institute, Ann Arbor, MI, United States

Nursing homes in the U.S have major staffing shortages that can lead to poor quality of care for residents. New minimum staffing standards mandated by the U.S. Centers for Medicare and Medicaid Services are intended to address this problem. The standards include requirements for a daily minimum of 2.45 hours of nurse aide (NA) time per resident (referred to as hours per resident day, or HPRD) and 0.55 daily hours of time by registered nurses (RNs) per resident. Most nursing homes in the U.S. do not currently meet these new standards, but there is significant variation among nursing homes around these overall HPRD levels. This has led us to investigate the following question: What characteristics of nursing homes are associated with a higher probability of meeting the newly proposed standards? Using logistic regression and national datasets, we investigate the association between whether a nursing home meets the standards and the following characteristics of the home:

- Whether the home is in a rural rather than non-rural county
- Whether the home is operated by a for-profit rather than nonprofit provider
- The percent of residents in the home who are persons of color
- The percent of residents in the home whose primary funding is Medicaid, as opposed to higher funding levels from Medicare or private payers

We also examine variation among states in the extent to which the above four characteristics affect the percent of nursing homes meeting the standards.

WB22

Summit - 342

Solutions for Low-Carbon Energy Transition

Contributed Session

Chair: Honghong Kou, Simon Fraser University, No. 266, Burnaby, BC, 710126, Canada

1 - Low-Carbon Optimization of User-Side Shared Energy Storage Based on Deep Reinforcement Learning

Tong Ye, Guangzhou Institute of Energy Conversion, Guangzhou, China, People's Republic of, Yuping Huang

Facing the challenges of global climate change and the need for improved energy efficiency, it is crucial to explore optimization strategies that reduce distribution network costs and carbon emissions. This study proposes a joint low-carbon optimization model for shared energy storage operators and distribution networks, considering tiered carbon trading. The model employs the Deep Deterministic Policy Gradient (DDPG) algorithm to achieve both economic benefits and environmental sustainability. Given the current constraints on carbon emissions and the instability of renewable energy supplies, the study first introduces an optimization framework that minimizes operational costs and carbon emissions through shared energy storage and all-day carbon trading. The scheduling problem is then formulated as a Markov Decision Process, defining the system's observation states, scheduling actions, and reward functions. Finally, the DDPG algorithm is applied for low-carbon economic dispatch.

2 - Coordinating TSOs, DSOs, and DERs Through a Tri-Level Decentralized Transactive Energy Market Framework

Tolga Karabas, Middle East Technical University, Ankara, Turkey, Meltem Peker Sarhan, Mustafa Kemal Tural

The increasing presence of Distributed Energy Resources (DERs) in power grids challenges the traditional unidirectional energy systems as they transition to bidirectional power flow. The inherent uncertainty of renewable DERs further complicates energy management. Despite these challenges, DERs provide valuable energy and flexibility services to power grids. This study introduces a novel tri-level decentralized transactive energy market framework to address these complexities. Within this framework, Distribution System Operators (DSOs) play a pivotal role as coordinators of local energy markets at the distribution level, serving as intermediaries to facilitate communication between Transmission System Operators (TSOs) and DERs using network-aware pricing and transactive control mechanisms. By enabling DER participation through optimized bidding strategies in local energy markets (and indirectly in wholesale markets via DSOs), the framework aims to optimize system operations while considering uncertainty through distributionally robust optimization methods. This approach offers a promising pathway to enhance grid security and reliability by addressing distributionally robust unit commitment and optimal power flow problems. By accounting for the dynamic interactions among TSOs, DSOs, and DERs, the proposed coordination approach seeks to establish a more resilient and efficient energy system management with optimized grid operations.

3 - Market Or Government Regulation: Impact On Operational Decarbonization

Eyup Emre Berk, Bilkent University, Ankara, Turkey

Operational decarbonization is the optimization of operational decisions in supply networks in the presence of carbon emissions penalties. We re-visit the stylized single location deterministic demand model with exogeneous carbon prices. In particular, we study the impact of carbon taxes imposed through government regulations and carbon pricing by consumers in the market on operational decisions when decision makers have objectives other than the traditional profit maximization motive. Our findings focus on the efficacy of carbon pricing mechanisms in reducing carbon footprints and have societal as well as potential regulatory implications.

4 - Asymmetric risk spillover network between energy stock index and China sector stock index based on energy transition perspective

Honghong Kou, Xidian university, Xi'an, China, People's Republic of

In the context of energy transition, elucidating the dynamic interactions and risk spillovers between energy markets and various industry stock indices is crucial for effective risk management and investment decisions. This paper employs the asymmetric TVP-VAR-DY model and MS-VAR model to construct an asymmetric time-varying energy-financial cross-market risk spillover network framework. This framework aims to describe asymmetric risk spillover relationships among 11 industry stock indices including traditional energy, clean energy, and Chinese energy, finance, and real estate. The research findings indicate: 1. Strong linkages exist between energy and Chinese industry stock index markets, with materials, industrial, and consumer discretionary sectors being the main sources of market risk, with the energy market primarily acting as a net receiver. 2. During bull markets, industry indices exhibit stronger predictability within the systemic network, while during bear markets, the energy market demonstrates greater predictability. 3. Further validation is provided on the association between extreme risk spillovers and significant economic events, policy interventions, investor sentiment, and global events. 4. Risk from Brent crude oil transmits to natural gas and coal, with Brent crude oil being more susceptible to contamination from clean energy than traditional energy. 5. With the advancement of energy transition policies, the spillover relationships exhibit a distinct three-stage transition, with risk gradually shifting towards new energy assets. Our research provides empirical evidence for the asymmetric interdependence and risk transmission mechanisms between energy and industry stock markets, revealing the impact of China's energy transition policies on the energy financial system.

5 - Understanding Sustainable Aviation Fuel Incentives with Low Carbon Fuel Policies: Market Designs in California

Adam Christensen, GAMS Development Corp, Fairfax, VA, United States, Colin Murphy, Julie Witcover

The US Environmental Protection Agency estimates that carbon emissions from the aviation sector was 8% of US transportation emissions and 2% of total US emissions in 2021; demand for these fuels is projected to increase dramatically. Decarbonizing the aviation sector is a grand challenge because commercial flight will be reliant on liquid fuels for the foreseeable future. Aviation fuel can be produced by hydrotreating lipids from agricultural/waste sources; other chemical pathways are also being explored. These fuels are often referred to as sustainable aviation fuels, or SAFs. SAFs have emerged as an attractive jet fuel alternative due to their lower embedded carbon content. However, the production of these fuels is limited; US aviation fuel consumption was ~23 billion gallons in 2021, but only 5 million gallons of SAF were consumed. California’s Low Carbon Fuel Standard (CA-LCFS) incentivizes these fuels by allowing producers to mint carbon credits and trade them in a marketplace. A wide variety of fuels can generate compliance credits making it difficult to model future credit market behavior. California is considering amendments to the LCFS that would incentivize SAFs although the exact incentive mechanism has not been determined. In this work we present our analysis of different market structures for integrating aviation fuels into the LCFS using a multi-agent partial equilibrium model of the California transport fuel sector and offer some policy-focused insights. This model is capable of resolving the different compliance strategies and endogenously determining the value of critical market incentives – specifically the carbon price.

WB23

Summit - 343

Advancements in Neural Network Applications

Contributed Session

Chair: Bingyi Su, North Carolina State University, Raleigh, NC, United States

1 - Optimization of Artificial Neural Networks in the Recognition of User Trust in Automation Utilizing Gaze Data Metrics

John Coughlin, Louisiana Tech University, Ruston, LA, United States, Mary Fendley

The prevalence of automated systems in man-machine systems necessitates the ability to effectively monitor user trust in those systems. The objective of this study was to create an Artificial Neural Network that can recognize the trust of an individual through the input of eye gaze data. Three models of Artificial Neural Networks were trained and tested on datasets gathered from previous studies. The procedure consisted of training an Artificial Neural Network with gaze data metrics as inputs and trust as output, utilizing the testing set as it is being trained to verify it is learning, and finally, utilizing the evaluation set to verify that it is applicable to unknown data. First, statistics regarding the best performances on the testing set were gathered. Second, with the best number of nodes in the hidden layer for each model now known, the best performance on both the training and testing sets were found for each of them. Finally, each model was evaluated on unknown data. Ultimately, the hybrid model ended up performing the best on known data, but the ReLU model ended up doing better overall on both known and unknown data.

2 - Convolutional Propagation on Graph Neural Networks for Hierarchical Network

Hyunjung Shin, Ajou University, Suwon, Korea, Republic of, yoonshin kwak, Junwoo Choi

Recently, Graph Convolutional Networks(GCN) have achieved success in various domains such as gene networks in bioinformatics, community detection in SNS, drug design in the pharmaceutical industry, and intrusion detection in the security field, etc. However, its application is limited only for a plane network or graph. If a network has hierarchical or multi-layered structure, it cannot be applicable to those networks, for instance, the omics networks layered by genome, proteome, diseasome, the financial networks layered by individual stocks, financial derivatives, global economical indicators, and the genealogy networks stratified by ancestor-descendant relationships. Another limitation of GCN is that its convolution process of sharing information with neighboring nodes is not flexible and tightly bounded with the GCN structure. Including more neighboring nodes requires more convolutional layers, which requires expensive computational resources. To cope with these limitations, we propose Graph Convolutional Smoothing Networks. In the proposed method, the GCN is redesigned to be expandable to hierarchical networks, and the convolution process employes the propagation process that adjusts the range of neighboring nodes through the smoothness parameter. Experiments show that the proposed method outperforms the comparison methods by taking advantage of its applicability to hierarchical networks and its flexibility of convolutional propagation.

3 - Enhancing Representation Ability of CNN via Color Inversion Channel

Dohyun Bu, Sungkyunkwan University, Suwon, Korea, Republic of, Bonyoung Koo, Chanho Kim, Gaeun Kim, Minjun Jeong, Yulim So, Jong-Seok Lee

Convolutional Neural Networks (CNNs) have proven to be potent tools for extracting features in various image processing tasks. However, traditional convolutional operations struggle to extract meaningful information from zero-valued pixels, limiting their utility in domains such as medical image analysis where such pixels are important. To address this challenge, we introduce a novel architecture based on dual-channel CNNs. We integrate both original channel and color-inversion channel CNNs, trained jointly to capture the distinct characteristics of each channel. Moreover, our approach exploits the orthogonality between extracted features, facilitating the learning of unique information from each channel. We demonstrate the effectiveness of our method in classifying medical images. Experimental results indicate that our approach can seamlessly integrate with pre-trained CNNs and surpass the state-of-the-art models in classification performance.

4 - Model-agnostic post-hoc explanation for image anomaly detection

Yulim So, Sungkyunkwan University, Suwon, Korea, Republic of, Seokho Kang

Image anomaly detection is a task of training anomaly detectors on normal images to identify whether a new image is anomalous. There have been significant research efforts focused on improving the accuracy of anomaly detectors. However, most of these anomaly detectors do not provide explanations for why certain images are identified as anomalous due to their “black-box” nature. In this study, we propose a surrogate explainer to provide explanations for the decisions made by any anomaly detector. When a query image is identified as an anomaly by a certain anomaly detector, the surrogate explainer is constructed using both normal images and the anomalous query image. Specifically, a convolutional neural network (CNN) is trained with a learning objective consisting of (1) classification loss to distinguish the query image from the normal images; (2) equivariant regularization loss to ensure equivariance between the query image and its attention map under affine transformations; and (3) complementary guided attention loss to expand the attention on the query image while suppressing the attention on the normal images. After training, the visual explanation of which pixels in the query image contribute to the anomaly is derived by applying gradient-weighted class activation map (Grad-CAM) to the CNN. Through experiments using the MVTec-AD benchmark, we demonstrate that our method provides explanations comparable to those of existing explainable anomaly detectors.

5 - Privacy-preserving federated learning using CNN-LSTM for activity recognition in manual material handling

Bingyi Su, North Carolina State University, Raleigh, NC, United States, Lu Lu, Liwei Qing, Sehee Jung, Xu Xu

Work-related musculoskeletal disorders (MSDs) are a major public health issue, affecting around 1.71 billion people worldwide. These conditions are often caused by manual material handling (MMH) tasks, making activity recognition in MMH crucial for preventing MSDs. However, centralized methods for training activity classifiers pose data privacy risks due to the sensitive nature of human data. Therefore, this study introduces a privacy-preserving federated learning framework using CNN-LSTM for activity recognition in MMH. Specifically, CNNs are utilized for their ability to extract spatial information, while LSTMs are employed for their effectiveness with time sequence data. Human motion data for training and test were collected using a motion tracking system from 11 participants. The body motions include commonly seen activities at the workplace, such as walking, lifting, lowering, stepping up, and reaching. The data was down-sampled to 15 Hz, simplified to 17 coordinates, normalized, and segmented into 2.67-second sequences. The sequences were then randomly split into training and test sets with an 8:2 ratio. For comparison purposes, both centralized and federated learning setup were applied. The centralized model achieved a precision of 0.984, recall of 0.983, and F1-score of 0.983. The federated model achieved a precision of 0.965, recall of 0.954, and F1-score of 0.958. These results show that federated learning offers comparable accuracy to centralized methods while ensuring data privacy, making it valuable for decision management in industrial settings to enhance workplace well-being.

WB24

Summit - 344

Advancements in Transportation and Logistics Optimization

Contributed Session

Chair: Rong Cheng, Eindhoven University of Technology, Eindhoven, 5612HZ, Netherlands

1 - Optimization of schedule-based barge timetable for a bi-modal urban logistics network

Rong Cheng, Eindhoven University of Technology, Eindhoven, Netherlands, Layla Martin, Claudia Archetti

Large facilities in city centers have high transport demands but operate in restricted space, resulting in low loading capacity and, consequently, long waiting times during peak hours, and negatively impacting traffic congestion and the environment. Some cities, such as Hamburg, can shift transportation to waterways thanks to an extensive network of canals. The barge timetable becomes crucial for freight owners' delivery decisions. This study develops a bi-level programming model for the barge timetabling problem in a bi-modal (truck delivery and truck-barge delivery) transport context, aiming to understand levers to incentivize freight owners to choose waterways using a game-theoretic analysis. The upper-level problem determines the barge timetable to maximize the share of truck-barge delivery. The lower-level problem describes freight owners' choice behavior by the user equilibrium, in which each freight owner chooses its delivery mode that minimizes its generalized cost according to the barge timetable of the upper-level model. The congestion effect at the facility's dock is modeled based on queueing theory. An iterative method is developed to solve the bi-level programming model. The iterative method includes: 1) an equilibrium assignment method to decide the delivery mode assignment. The equilibrium assignment method is embedded with a pointwise stationary fluid approximation method to evaluate the queueing system. 2) a branch-and-bound algorithm for the barge timetable design. Experiments are conducted to illustrate the model's properties and the performance of the solution method.

2 - Two-echelon multi-trip vehicle routing problems with synchronization under delays

Cigdem Karademir, Delft University of Technology, Delft, Netherlands, Breno Beirigo, Bilge Atasoy

Meeting the increasing demand in cities has become challenging under the regulations limiting freight vehicles to improve mobility and reduce carbon footprints. By effectively coordinating with larger vehicles, new electric vehicle technologies can overcome restrictions in driving range and storage capacity. This study proposes integrated water- and land-based transportation systems using light electric freight vehicles and autonomous vessels as mobile depots to enhance the efficiency of city logistics. However, such systems introduce dependency between different vehicles at the transshipment locations and face potential delays that can impact the reliability of the entire service network.

Dedicating storage at these locations might mitigate the impact of delays but limit the reliability to location and capacity decisions at the strategic level. On the other hand, public spaces such as parking spots or transportation stops can be used as on-demand transshipment points, enabling revisiting these decisions at the operational level. However, this flexibility requires spatio-temporal synchronization between interacting vehicles and might propagate the delays in the system.

This study provides comparative analyses of the impact of delays on integrated service designs with and without storage, 2-echelon-S and 2-echelon-F, respectively. We model these systems as two-echelon multi-trip vehicle routing problems with synchronization, addressing delays through a two-stage stochastic program with simple recourse. Preliminary experiments on small networks show that 2-echelon-F is better at reducing expected lateness and offers greater flexibility in logistics re-organization to minimize delays' impact. To further analyze the reliability of such sustainable logistics solutions in cities, different delay sources will be studied.

3 - Wattage vs. Mileage: Unpacking the Dilemma of Electric Long-Haul Trucking

David Menachof, Florida Atlantic University, Boca Raton, FL, United States, Nihat Ahmed

Electric long-haul semi-trucks are feasible for about 72% of origin-destination (O-D) pair combinations using existing truck stops (if enhanced with charging stations). Just because they are feasible, does not mean they are practical. While current charging infrastructure lacks the number of mega-chargers required to make the long-haul electric trucking practical, even if the charging network is expanded by a factor of 10, it would still not be enough to make it a viable alternative to the diesel tractor unit. Examination of the Freight Analysis Framework will show that realistic volumes of electric semi-truck units on any O-D combination will be a small percentage of total volume on each route.

Using QGIS, the US major highway network will be used to show the total volume over any segment of road. In some cases, multiple OD combinations will use a specific segment. Combining existing charging station locations along with projected locations, we will project potential daily volumes that can be supported. Using actual data over a segment compared to projected charging volumes, we will determine projected shortfall of charging capacity. We will examine maximum volumes over an OD route pair based on the distance and charging stops required. Charging capacity available at each stop will be calculated. The overall capacity constraint for the entire route will be determined by the lowest charging capacity among all stops along a route.

If the charging infrastructure capacity problem cannot be solved, then where does that leave the future of electric long haul semi-trucks?

4 - Enhancing Human Planning Decisions with AI: An Experimental Investigation and Evidence

Lijia Tan, Eindhoven University of Technology, Eindhoven, Netherlands, Eirini Spiliotopoulou, Willem van Jaarsveld

We experimentally investigate the impact of AI plans on human planning decisions in supply chains. The AI plans are based on a neural network and reinforcement learning algorithm, which provide recommendations for all decisions that humans need to make. We conducted two experiments. The first experiment examines the impact of AI plans on inexperienced and experienced planners, respectively. The first experiment shows that the presence of AI plans effectively improves the performance of inexperienced planners but barely affects experienced planners. Specifically, such AI plans assist inexperienced planners in performing as well as experienced planners, but they do not improve experienced planners' performance. The second experiment examines the impact of AI plans under a time-pressure environment. The results show that the presence of AI is more effective regardless of human experience when time pressure is present. In this study, our discussion on the interactive effects between AI plans and human experience, and between AI plans and time pressure, provides insights into the applications of AI in future supply chains.

WB25

Summit - 345

Emerging Topics for Electric Vehicles

Contributed Session

Chair: Yamil Essus, North Carolina State University, Raleigh, NC, United States

1 - Estimating Access Risk Due to Vehicle Electrification During Blackouts in North Carolina

Yamil Essus, North Carolina State University, Raleigh, NC, United States, Benjamin Rachunok

Vehicle electrification is a key component of sustainable development goals, yet the mass adoption of electric vehicles can lead to unintended consequences for community mobility during natural hazards. Electric vehicles pose a challenge to owners during natural hazards that lead to power outages, as a lack of home charging limits the mobility of EV owners. Due to the direct relation between driving distance to essential services and vehicle battery consumption, changes in mobility will be impacted by geographic and technological factors. Geography determines the driving distance to essential services, which will be translated to electricity consumption devoted to transportation, and technology determines the size of electric vehicle batteries and vehicle efficiency. The linkage between mobility, electric power availability and quality of life has broad implications for community resilience as equitable access to essential services has been identified as the most important aspect of community resilience. We present the results of a computational framework for estimating the risk of losing access to essential services during a prolonged blackout in North Carolina, assuming complete adoption of electric vehicles. We leverage an open-source routing engine to compute driving distance to services, census data to identify driving requirements (such as school enrollment and commuting by private car) and a database of previous blackouts to estimate county-level blackout risk. Additionally, we incorporate a heuristic to approximate distance driven on commuting using isochrone curves. Finally, we explore the sensitivity of our results to vehicle battery capacity, which is tightly linked to vehicle affordability.

2 - Optimization of electric truck charging infrastructure considering GHG emissions

WEILIANG DENG, University of Arizona, Tucson, AZ, United States

Compared with widely adopted passenger Electric Vehicles (EVs), electric trucks (e-trucks) have high requirement for charging infrastructure and low operational efficiency (such as long charging time and frequent charging stops), which hinders market adoption. A reliable and easily accessible charging infrastructure network can facilitate this transformation from traditional truck fleets to electric alternatives. This paper proposes a charging infrastructure location optimization strategy to develop such charging stations network for e-trucks. Candidate charging station locations of e-trucks are identified using ArcGIS and evaluated based on various factors such as accessibility, grid network, power supply, and corridors priority. Besides, the source of electricity used to refill e-trucks' batteries will impact the overall GHG emissions during the operation stage of e-trucks. To ensure the lowest overall GHG emissions of e-trucks, energy mix of power supplier is incorporated to evaluate the overall environmental impact. A mixed integer programming model is formulated to minimize the total economic investment and GHG emissions. The model is then applied to investigate operation of commercial e-truck fleet in Arizona. Other managerial insights are further discussed.

3 - Optimizing Perimeter Control for Mixed Traffic of Conventional and Electric Vehicles in Large-Scale Transportation Networks

Ehsan Kamjoo, Michigan State University, East Lansing, MI, United States, Fatemeh Fakhrmoosavi, Ali Zockaie

Electric Vehicle (EV) technology offers an effective solution to mitigate traffic-induced air pollution in urban areas. However, the slow projected increase in EV market share leads to a prolonged period of mixed traffic of EVs and Conventional Vehicles (CVs). Given the non-uniform distribution of congestion and its related environmental issues across the network, there is a need for innovative traffic management strategies that prioritize EVs in areas experiencing high levels of congestion and pollution, such as Central Business Districts. This study introduces an innovative control strategy that dedicates specific links on the periphery of network regions to EVs to minimize both total travel time of vehicles as well as vehicular emissions over the network. Unlike existing studies that do not differentiate between CVs and EVs in perimeter control, this approach manages the accumulations of each vehicle type in each region by adjusting the traffic signals at the boundaries. A Model Predictive Control (MPC) method is employed to solve the optimization problem, coupled with a mesoscopic traffic simulation tool, DYNASMART-P, to simulate the movement of vehicles throughout the large-scale network. The proposed framework is successfully applied to the Chicago network, evaluating congestion and emissions under scenarios with and without perimeter control, including scenarios that do not differentiate between CVs and EVs. Different scenarios for the market share of EVs are also considered to evaluate their impact on the performance of perimeter control. The results provide valuable insights for policymakers on the effective implementation of this strategy in presence of EVs.

4 - Decarbonization Strategies: Balancing Investments in Renewables, Electric Vehicles, and Carbon Capture

Daniel Fontecha, KAUST, Thuwal, Saudi Arabia, Ricardo M. Lima, Omar Knio, Omar Knio

To which extent should governments invest in renewables and electric vehicles (EVs)? Renewables variability and significant investment costs in batteries and transmission hamper the decarbonization of the power sector and thus penetration of EVs.

EVs add extra load to the power sector, but can also support it through load-shifting and vehicle-to-grid (V2G), in addition to decreasing the carbon emissions of the transportation sector.

Therefore, given the interactions between power, transportation, and carbon emissions, it is crucial to quantify how long-term investments in renewables, EVs and carbon-capture technologies mutually affect each other to identify the most efficient path for decarbonizing our economies.

To address these questions, a capacity expansion problem is formulated as stochastic mixed-integer program, minimizing the total cost of a power system linked with the road transportation sector while adhering to annual emission limits. A two-stage decision framework is adopted to address uncertainty during the operational stage. The model is applied to Saudi Arabia, utilizing an extensive formulation and a parallelized solution strategy for multiple instances.

The study reveals a positive net effect of EVs on the power sector, with benefits magnifying as the share of renewable generation increases. For instance, an uncontrolled EV penetration of 10% reduces the total cost of a power sector under tight emission constraints by up to 17%. The optimal renewable share increased from 56% to 63%. However, the whole panorama changes when carbon capture technologies are incorporated in the model, demonstrating the importance of evaluating simultaneously all available technologies to trace decarbonization pathways.

5 - An Analysis of Capacity and Pricing Decisions for Renewable Energy-Powered EV Charging Stations on Highways

Zhuo Feng, Dalian University of Technology, Dalian, China, People's Republic of, Xin Chen, Ying Gao, Shuibo Zhang

The increasing adoption of electric vehicles (EVs) has led to the installation of solar panels to provide renewable energy to EV charging stations on highways, contributing to a decrease in carbon emissions associated with EV power consumption. This paper presents a game-theoretic model to study the capacity decisions (i.e., the capacities of solar panels and EV charging stations) and pricing decisions (i.e., tolls and charging prices) of a private firm. Additionally, we examine the government's necessary subsidy levels to meet predefined carbon emission reduction targets.

WB26

Summit - 346

Innovative Approaches in Financial Market Analysis and Forecasting

Contributed Session

Chair: Grace Babalola, Binghamton University (SUNY), Johnson City, NY, United States

1 - A Network Approach for the Financial Market Regimes: A Study of Chinese Stock Market

Xiaodi Zhu, University of Wisconsin-La Crosse, La Crosse, WI, United States, Anqi Liu

In the investment world, identifying, measuring and modelling connections among different assets are important issues in risk management. Beyond portfolio management, when viewing the financial market as a complex system, connections among assets can not only reflect stability of the entire financial system, but also present the vulnerability, contributions, and spillovers of each asset. This study uses network method to study the Chinese stock market change in the past 10 years. Using a automatic regime detection method, this study detects four major financial regimes in the market, and visualize their structure change.

2 - Spatial Competition with Online Platforms: An Empirical Analysis of the Wealth Management Product Market

Shuoxun Zhang, Sichuan University, Chengdu, China, People's Republic of, Chao Ma

Previous empirical studies have analyzed multiple sectors and found that if different, most time online prices are lower than offline. We study the market of a financial product, the wealth management product (WMP), which is the largest component of China’s shadow banking sector. We find that prices of WMPs sold online is significantly higher than those sold at banks’ physical branches. The explanation is that customers need to pay travel costs for visiting physical branches and thus banks have to charge lower prices at physical branches to attract customers from online platforms.

The WMP market provides an ideal scenario to identify the effect of firms’ spatial competition that contributes to the gap between online and offline prices. The reason is that online and offline channels for WMP issuing are not differential in other dimensions that could affect prices, including delivery costs, search costs, convenience to inspect product qualities, and waiting costs.

We also find that a bank’s online-offline price gaps decrease with both the number of incumbent banks and the bank’s branch density in the city, although the former tends to decrease a bank’s market power whereas the latter tends to increase a bank’s market power. As the number of competitors increases, a bank can still maintain some local market power in the offline market over residents near its physical branches but cannot do so in the online market. In contrast, a bank’s branch density in a city directly increases the bank’s local market power only in the offline market.

3 - Dynamic Risk Budgeting Portfolio using Value-at-Risk and Expected Shortfall

Wen-Yi Lee, National Taipei University of Business, Taipei, Taiwan, Ya-Hsin Chou

Risk budgeting portfolio studies are a crucial aspect of asset allocation, aiming to incorporate different risk measurements. However, these studies need to pay more attention to the continuous nature of risk over time. We propose a risk budgeting portfolio that continuously detects Value-at-Risk (VaR) and expected shortfalls to address this issue. In this study, we present two models that consider different trading timings.

The first model is the portfolio model with static trading timing. This model involves rebalancing portfolios at fixed intervals, resulting in a predetermined rebalancing schedule with portfolios adjusted at predetermined intervals. While this approach offers a reliable method for risk budgeting, it may not account for market conditions that arise between these predetermined intervals. On the other hand, the second model utilizes dynamic trading timing. This approach involves trading portfolios when their VaR or expected shortfall exceeds critical values, allowing for a more nuanced and efficient response to real-time risk based on market conditions. This model offers more comprehensive risk budgeting portfolio studies that account for the continuous nature of risk and offer both static and dynamic trading strategies. Overall, our proposed models provide a more comprehensive approach to risk budgeting portfolio studies that accounts for the continuous nature of risk and offers both static and dynamic trading strategies.

4 - The Value of Historical Financial Statements in Predicting the Innovation Performance of Technology-Based SMEs: A Machine Learning Approach

bin ke, National University of Singapore, Singapore, Singapore, Xinjie Ma, Yanbo Wang

Utilizing a proprietary dataset of grant applications from technology-based SMEs to China's Innovation Fund for Small Technology-based Firms (Innofund), this study evaluates the utility of historical financial statements for assessing SMEs' innovation potential. Our findings demonstrate that an advanced XGBoost model, using 55 financial statement data items only, predicts innovation performance more effectively than an XGBoost model based on human experts' evaluation scores of the entire grant application. Additionally, merging financial data with expert scores does not improve prediction accuracy, indicating no incremental value from expert discretion. Analysis of Shapley values highlights distinct evaluation criteria differences between human experts and the XGBoost model. Our results demonstrate the promise of leveraging advanced machine learning with historical financial data in aiding government grant agencies to identify SMEs with high innovation potential.

5 - Analysis of Stock Market Performance During Election Year and Non-election Year Using Differential Entropy and Mutual Information

Grace Babalola, Binghamton University (SUNY), Binghamton, NY, United States, Ademola Adedoyin, Tejaswini Dinesh, Duxiao Hao, Uthman Kareem, Hamid Kolawole, Bukunmi Ofili, Simeon Ogunbunmi

Stock market volatility is one of the key factors in deciding the potential investment in the stock market. The volatility in the stock market in the months preceding an election has been observed for the past few decades. Several studies show that there has been a significant increase in the market value in the third year of a presidential term similarly to the year 2019 observation. However, the 'Presidential Election Theory' used in most of these studies is susceptible to other unpredictable factors which have been observed during the current year. Thus, this study attempts to evaluate a distinction in the performance of the stock market during election (X) and non-election (Y) years considering other susceptible factors not encountered in other studies, to give an overview of the distinction in the risks associated with investing. This study analyzed the Dow Jones Industrial Average Annual Stock Price using differential entropy and mutual information. The differential entropy calculated was 0.0848001 with $X\mu$:11,232.2 (X_{SD}: 7,625.58) and $Y\mu$:10,001(Y_{SD}:7,190.27), while the mutual information was calculated to validate the differential entropy value. Using the mutual information formula HX + HY – HXY, the output value is 5.3529. Invariably concluding based on the differential entropy that although the uncertainty in the information shared in the stock market performance during election and non-election year is approximately 5.35 bits of information, the stock market performance is independent of the year of election. Hence, one may ignore the condition of election year while deciding to invest in the stock market.

WB27

Summit - 347

Advances in Sequential Decision-Making with E-Commerce Applications

Invited Session

Revenue Management and Pricing

Chair: Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States

1 - Primal-Dual Algorithm for Online Resource Allocation without Resolving

Yehua Wei, Duke University, Durham, NC, United States, Jiaming Xu, Sophie Yu

We present a simple primal-dual algorithm for online resource allocation problems with finite query types. We show that our algorithm achieves constant regret, i.e., a regret that is independent of T, when the fluid solution is non-degenerate. Unlike many other constant regret algorithms in the literature, our algorithm does not require the fluid problem to be resolved over time, making it more computationally efficient in large-scale resource allocation. In addition, we discuss our regret's dependency on system parameters outside of T, and how it compares with other algorithms in the literature.

2 - Liquid Welfare guarantees for No-Regret Learning in Sequential Budgeted Auctions

Giannis Fikioris, Cornell University, Ithaca, NY, United States, Eva Tardos

We study the liquid welfare in sequential first-price auctions with budgeted buyers. We use a behavioral model for the buyers, assuming a learning style guarantee: the utility of each buyer is within a γ factor ($\gamma\geq 1$) of the utility achievable by shading her value with the same factor at each iteration. We show a $\gamma+1/2+O(1/\gamma)$ price of anarchy for liquid welfare when valuations are additive. This is in stark contrast to sequential second-price auctions, where the resulting liquid welfare can be arbitrarily smaller than the maximum liquid welfare, even when $\gamma=1$. We prove a lower bound of γ on the liquid welfare loss under the above assumption in first-price auctions.
Our liquid welfare results extend when buyers have submodular valuations over the set of items they win across iterations with a slightly worse price of anarchy bound of $\gamma+1+O(1/\gamma)$ compared to the guarantee for the additive case.

3 - Optimizing Inventory Placement for a Downstream Online Matching Problem

Boris Epstein, Columbia University, New York, NY, United States, Will Ma

We study the inventory placement problem of splitting Q units of a single item across warehouses, in advance of a downstream online matching problem that represents the dynamic fulfillment decisions of an e-commerce retailer. This is a challenging problem both in theory, because the downstream matching problem itself is computationally hard, and in practice, because the fulfillment team is constantly updating its algorithm and the placement team cannot directly evaluate how a placement decision would perform.

We compare the performance of three placement procedures based on optimizing surrogate functions that have been studied and applied: Offline, Myopic, and Fluid placement. On the theory side, we show that optimizing inventory placement for the Offline surrogate leads to a $(1-(1-1/d)^d)/2$-approximation for the joint placement and fulfillment problem. We assume d is an upper bound on how many warehouses can serve any demand location and that stochastic arrivals satisfy either temporal or spatial independence. The crux of our theoretical contribution is to use randomized rounding to derive a tight $(1-(1-1/d)^d)$-approximation for the integer programming problem of optimizing the Offline surrogate. We use statistical learning to show that rounding after optimizing a sample-average Offline surrogate, which is necessary due to the exponentially-sized support, does indeed have vanishing loss.

On the experimental side, we extract real-world sequences of customer orders from publicly-available JD.com data and evaluate different combinations of placement and fulfillment procedures. Optimizing the Offline surrogate performs best overall, even compared to simulation procedures, corroborating our theory.

4 - Online Bin Packing with Load Balancing

Hanzhang Qin, National University of Singapore, Singapore, Singapore, Yan Chen, Zheng Tan, Yugang Yu

The balls-into-bins approach, extensively studied in Computer Science, can address load balancing against various uncertain environments. However, in real-world applications such as cloud resource allocation and last-mile delivery, constraints on bin capacity often necessitate more sophisticated approaches. In this paper, we present a novel framework centered on a level-based constrained optimization formulation aimed at simultaneously minimizing the surplus bins and mean-square deviation from the mean level in a lexicographically optimal manner. We introduce a new penalized Lagrangian function as the criterion for bin selection. Through theoretical analysis and numerical experiments, we establish the effectiveness of our proposed algorithm in achieving near-optimal solutions, particularly in sparse scenarios, when compared to offline optimal solutions.

WB28

Summit - 348

Pricing and Other Considerations in Business Operations

Contributed Session

Chair: Ruibing Wang, University of Mannheim, Schwetzingen, Germany

1 - Detecting Forward-looking Statements Fraud in Earnings Conference Calls Using Graph Attention Network Approach

Elaheh Aghazadeh, University of Massachusetts Lowell, Lowell, MA, United States, Julie Zhang

The detection of forward-looking statement fraud in earnings conference calls is essential for maintaining the integrity of financial markets due to the inherent uncertainty of predictions about a company's future financial performance and strategic direction. These statements, which often include earnings forecasts and assessments of market potential, are particularly vulnerable to manipulation. Whether these manipulations are overly optimistic or pessimistic, they can significantly alter market perceptions, mislead investors, and detrimentally impact stock prices. Recognizing the critical need for robust detection mechanisms, this study employs a Graph Attention Network (GAT) approach. This method leverages advanced machine learning techniques to analyze complex relational data from earnings calls, aiming to identify deceptive patterns effectively. The objective is to enhance the capability of regulators and investors to verify the credibility of corporate disclosures, thereby ensuring market fairness and stability. The paper uses the GAT methodology and investigates its transformative potential in financial fraud detection, thus supporting more informed regulatory and investment decisions.

2 - Scoping due diligence process(es) in private equity transactions

Benoît Chevalier-Roignant, emlyon business school, Ecully, France

To make the right investment decisions, a private equity (PE) firm typically gathers critical information about the would-be acquiree (``target") as part of due diligence processes. Such information acquisition often requires the hiring of consultants who carry out various due diligence streams. This paper develops a simple mathematical model to study the optimal scope of one's due diligence process, first examining when to engage a consulting firm to conduct a particular due diligence stream (e.g., a financial or a commercial due diligence) and then exploring whether (and if so, how) the PE fund adjusts the project scopes if several consultants work on different streams in parallel. We identify key factors that drive the PE fund's decision and obtain useful managerial insights.

(Co-authored by Tiziano De Angelis, University of Torino)

3 - Assortment Planning and Pricing with Consumer Searching: The Role of Anticipated Regret

Ruibing Wang, University of Mannheim, Mannheim, Germany, Cornelia Schön, Oliver Vetter

Consumers often search for product information to resolve valuation uncertainties before purchasing. Since they cannot examine all alternatives because of their limited information-acquisition ability, they conduct a two-stage search, called consider-then-choose. In the first stage, they decide a consideration set which is a subset of all available alternatives. In the second stage, they resolve the uncertainty about all products in this consideration set and choose the one with the highest utility. We extend assortment optimization under consumer choice behavior by incorporating consumer anticipated regret into the consideration set formation. Specifically, after purchase, when consumers passively notice whether an unconsidered product is more suitable than the purchased one or not, they might regret having decided a small consideration set and not devoted enough effort to searching or regret having decided a large consideration set and devoted too much effort to searching. We show that if consumers anticipate the post-purchase regret when deciding a consideration set, in some conditions the firms should offer some products that consumers will not consider. Based on this finding, we develop fully-polynomial approximation schemes or exact conic formulation for a variety of assortment problems under the consider-then-choose models with anticipated regret. For the joint assortment planning and pricing problem with homogeneous consumers, we show that the intrinsic-utility ordered assortment and the price policy that charges the same price for the products except at most those outside the consideration set and one within it are optimal.

WB29

Summit - 420

Approximation Algorithms

Invited Session

OPT: Integer and Discrete Optimization

Chair: Heather Newman, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Rudy Zhou, Carnegie Mellon University, Pittsburgh, PA

1 - Identifying the minimum value under explorable stochastic uncertainty

Hessa Al-Thani, University of Michigan, Ann Arbor, MI, United States, Viswanath Nagarajan

Consider a set of bounded random variables, each of which can be probed to determine its realized value. We are interested in determining the minimum (or maximum) value of these random variables upto some precision, while minimizing the number of probes. This is an abstract model that has applications in hiring and bid-selection, where one wants to identify an option having the maximum/minimum value (upto some precision)
by evaluating the fewest number of available options. We provide a 4-approximate policy for this problem. Interestingly, our policy is "non adaptive",
that is it probes the random variables in an a-priori fixed order until a suitable stopping rule. We note that an optimal policy may be highly adaptive, choosing the next variable to probe based on all prior realizations.

2 - Bayesian Stochastic Probing

Rudy Zhou, Carnegie Mellon University, Pittsburgh, PA, United States

We introduce a new stochastic probing problem with correlations across items to be probed. This problem has applications in stochastic knapsack with correlations across items and Bayesian active search. We give non-adaptive and adaptive approximation algorithms for this problem.

3 - Simultaneously Approximating All lp-norms in Correlation Clustering

Heather Newman, Carnegie Mellon University, Pittsburgh, PA, United States, Sami Davies, Benjamin Moseley

This paper considers correlation clustering on unweighted complete graphs. We give a combinatorial algorithm that returns a single clustering solution that is simultaneously O(1)-approximate for all ℓ_p-norms of the disagreement vector; in other words, a combinatorial O(1)-approximation of the all-norms objective for correlation clustering. This is the first proof that minimal sacrifice is needed in order to optimize different norms of the disagreement vector. In addition, our algorithm is the first combinatorial approximation algorithm for the ℓ_2-norm objective, and more generally the first combinatorial algorithm for the ℓ_p-norm objective when $1<p<\infty$. It is also faster than all previous algorithms that minimize the ℓ_p-norm of the disagreement vector, with run-time $O(n^{\omega})$, where $O(n^{\omega})$ is the time for matrix multiplication on $n\times n$ matrices. When the maximum positive degree in the graph is at most Δ, this can be improved to a run-time of $O(n\Delta^2 \log n)$.

4 - Fast Combinatorial Algorithms for Efficient Sortation

Madison Van Dyk, Amazon, Seattle, WA, United States, Jochen Koenemann

Modern parcel logistic networks are designed to ship demand between given origin, destination pairs of nodes in an underlying directed network. Efficiency dictates that volume needs to be consolidated at intermediate nodes. In practice, such consolidation requires parcel-sortation. In this work, we propose a mathematical model for the physical requirements, and limitations of parcel sortation. We then show that, in the general model, it is NP-hard to determine whether a feasible sortation plan exists. We discuss several special settings, where (near-)feasibility of a given sortation instance can be determined efficiently. The algorithms we propose are fast and build on combinatorial witness set type lower bounds that are reminiscent and extend those used in earlier work on degree-bounded spanning trees and arborescences.

WB30

Summit - 421

Verification of Decision-Making Algorithms

Invited Session

OPT: Linear and Conic Optimization

Chair: Vinit Ranjan, Princeton University, Princeton, NJ, United States

Co-Chair: Bartolomeo Stellato, Princeton University, Princeton, NJ, United States

1 - α,β-CROWN: A Linear Bound Propagation Framework for Formal Verification of Neural Networks

Huan Zhang, University of Illinois Urbana-Champaign, Urbana, IL, United States

Neural network (NN) verification seeks to rigorously guarantee some properties of neural networks, including safety, robustness, and correctness, which are essential for using NNs in mission-critical systems. It can be formulated as a large-scale non-convex optimization problem, but off-the-shelf solvers such as mixed integer programming (MIP) can hardly scale to NNs with practical sizes. To address this challenge, I will introduce a novel framework called “linear bound propagation methods”, which exploits the structure of this problem by efficiently propagating inequalities on neural networks. It is scalable to large neural networks with millions of neurons with GPU acceleration and achieves up to three orders of magnitude speedup compared to traditional algorithms. The series of successful bound propagation algorithms leads to the development of the award-winning α,β-CROWN toolbox (https://abcrown.org). Finally, I will introduce my recent work on applying α,β-CROWN to give formal stability and safety guarantees for neural network-enabled decision-making and nonlinear control problems.

2 - Computer-Assisted Design of Accelerated Composite Optimization Methods

Uijeong Jang, Seoul National University, Seoul, Korea, Republic of, Shuvomoy Das Gupta, Ernest Ryu

The accelerated composite optimization method FISTA (Beck, Teboulle 2009) is suboptimal, and we present a new method OptISTA that improves upon it by a factor of 2. The performance estimation problem (PEP) has recently been introduced as a new computer-assisted paradigm for designing optimal first-order methods, but the methodology was largely limited to unconstrained optimization with a single function. In this work, we present a novel double-function stepsize-optimization PEP methodology that poses the optimization over fixed-step first-order methods for composite optimization as a finite-dimensional nonconvex QCQP, which can be practically solved through spatial branch-and-bound algorithms, and use it to design the exact optimal method OptISTA for the composite optimization setup. We then establish the exact optimality of OptISTA with a novel lower-bound construction that extends the semi-interpolated zero-chain construction (Drori, Taylor 2022) to the double-function setup of composite optimization. By establishing exact optimality, our work concludes the search for the fastest first-order methods for the proximal, projected-gradient, and proximal-gradient setups.

3 - Verification of First-Order Methods for Parametric Quadratic Optimization

Vinit Ranjan, Princeton University, Princeton, NJ, United States, Bartolomeo Stellato

We introduce a numerical framework to verify the finite step convergence of first-order methods for parametric convex quadratic optimization. We formulate the verification problem as a mathematical optimization problem where we maximize a performance metric (e.g., fixed-point residual at the last iteration) subject to constraints representing proximal algorithm steps (e.g., linear system solutions, projections, or gradient steps). Our framework is highly modular because we encode a wide range of proximal algorithms as variations of two primitive steps: affine steps and element-wise maximum steps. Compared to standard convergence analysis and performance estimation techniques, we can explicitly quantify the effects of warm-starting by directly representing the sets where the initial iterates and parameters live. We show that the verification problem is NP-hard, and we construct strong semidefinite programming relaxations using various constraint tightening techniques. Numerical examples in nonnegative least squares, network utility maximization, Lasso, and optimal control show a significant reduction in pessimism of our framework compared to standard worst-case convergence analysis techniques.

4 - Data-Driven Performance Guarantees for Classical and Learned Optimizers

Bartolomeo Stellato, Princeton University, Princeton, NJ, United States, Rajiv Sambharya

We introduce a data-driven approach to analyze the performance of continuous optimization algorithms using generalization guarantees from statistical learning theory. We study classical and learned optimizers to solve families of parametric optimization problems. We build generalization guarantees for classical optimizers, using a sample convergence bound, and for learned optimizers, using the Probably Approximately Correct (PAC)-Bayes framework. To train learned optimizers, we use a gradient-based algorithm to directly minimize the PAC-Bayes upper bound. Numerical experiments in signal processing, control, and meta-learning showcase the ability of our framework to provide strong generalization guarantees for both classical and learned optimizers given a fixed budget of iterations. For classical optimizers, our bounds are much tighter than those that worst-case guarantees provide. For learned optimizers, our bounds outperform the empirical outcomes observed in their non-learned counterparts.

5 - Derivative-informed Abstraction of Neural Networks for Reachability Analysis

Mahyar Fazlyab, Johns Hopkins University, Baltimore, MD, United States

A current limitation of state-of-the-art verification methods for neural networks is their inability to efficiently handle continuously differentiable neural networks, as they are purposefully tailored to piece-wise-linear neural networks, and hence, become inherently inapplicable or at least highly inefficient for continuously differentiable neural networks. In this talk, I will discuss how we can leverage higher-order derivative information to drastically improve the scalability and computational efficiency of verifying neural networks for safety, robustness, or other desirable properties. More specifically, the proposed approach abstracts (i.e., over-approximates) the map of neural networks by higher-order Taylor series expansions and bounding the higher-order terms. In contrast to the existing verification methods, the resulting abstraction locally preserves the derivative information, thereby yielding accurate bounds for small uncertainties. We tailor this approach in adversarial robustness certification as well as reachability analysis.

WB31

Summit - 422

Advances in Robust and Risk-Sensitive Markov Decision Processes

Invited Session

OPT: Optimization Under Uncertainty

Chair: Jefferson Huang, Naval Postgraduate School, Monterey, CA, United States

1 - Stationary Optimal Policies in Risk-averse MDPs with EVaR

Marek Petrik, University of New Hampshire, Durham, NH, United States

The infinite-horizon discounted objective is popular in reinforcement learning in part because it admits stationary policies and convenient analysis. Unfortunately, all common discounted risk-averse objectives require history-dependent policies that must be computed using complex state augmentation. In this paper, we show that the total reward objective (also known as a stochastic shortest path) under the Entropic Risk Measure (ERM) and Entropic Value at Risk (EVaR) can be optimized by a stationary policy. In addition, the optimal policy can be computed using value iteration, policy iteration, and even linear programming. We establish our results under the relatively mild condition of transient MDPs allowing for both positive and negative rewards, unlike prior work. Our analytical and numerical results suggest that the total reward criterion may be preferable to the discounted criterion in a broad range of risk-averse reinforcement learning problems.

2 - Risk-Sensitive Discounted MDPs with Monotone Utilities and Unbounded Costs and Rewards

Eugene Feinberg, Stony Brook University, Stony Brook, NY, United States

This talk deals with optimization of generalized risk-sensitive functions which are modelled using increasing functionals of the discounted sums of the per-stage cost or rewards. Such functionals are defined on the full paths of the underlying random dynamical system represented by a controlled Markov chain on general state and action spaces. For the cost minimization case with nonnegative and possibly unbounded-above per-stage costs, we prove the validity of optimality equations and existence of optimal policies with noncompact action spaces, both for the nite and in nite horizon setup. For the reward maximization case with nonnegative and possibly unbounded-above per-stage rewards, we prove the validity of optimality equations and existence of optimal policies under a mild assumption of asymptotic uniform integrability of the value function, again both for the nite and in nite horizon setup. For both cost and reward cases, we also prove the convergence of the nite-horizon optimal value function to the in nite horizon optimal one. For the reward case, we also provide some new results on the connection between Lyapunov stability of the underlying transition kernel and the asymptotic uniform integrability of the infnite-horizon value function.

3 - Dynamically Maintaining Infrastructure Networks

Jefferson Huang, Naval Postgraduate School, Monterey , CA, United States

This talk is about Markov decision models of sequential maintenance planning problems in infrastructure networks, such as fuel or water distribution networks. We focus on cases where the one-step cost associated with each feasible state-action pair is the optimal objective function value of an associated network flow problem. In general, optimal policies for these models are complicated, and can quickly become difficult to compute as the size of the network increases. Moreover, the possibility of climate or adversary-driven disruptions motivates the search for policies that are robust. We describe recent work addressing these issues that also draws connections to the network interdiction literature.

WB32

Summit - 423

Recent Advances in Global Stochastic Optimization

Invited Session

OPT: Global Optimization

Chair: Burla Ondes, Purdue University, West Lafayette, IN, 47906, United States

1 - Rate-Optimal Budget Allocation for the Probability of Good Selection

David Eckman, Texas A&M University, College Station, TX, United States, Taeho Kim

This talk considers the allocation of simulation effort in a ranking-and-selection (R&S) problem with the goal of selecting a system whose performance is within a given tolerance of the best. We apply large-deviations theory to derive an optimal allocation for maximizing the rate at which the so-called probability of good selection (PGS) asymptotically approaches one, assuming that systems' output distributions are known. An interesting property of the optimal allocation is that some good systems may receive a sampling ratio of zero. We demonstrate through numerical experiments that this property leads to serious practical consequences, specifically when designing adaptive R&S algorithms. In particular, we observe that the convergence and even consistency of a simple plug-in algorithm designed for the PGS goal can be negatively impacted. We offer empirical evidence of these challenges and a preliminary exploration of a potential correction.

2 - Monte Carlo Partitioned Bayesian Optimization for Treed Black Box Functions

Mohit Malu, Arizona State University, Tempe, AZ, United States, Giulia Pedrielli, Gautam Dasarathy, Andreas Spanias

Bayesian Optimization (BO) frameworks typically assume the function to be optimized is stationary (homogeneous) over the domain. However, in many real-world applications, we often deal with functions that present a rate of variation across the input space. In this paper, we optimize heterogeneous functions where a finite set of homogeneous functions defined over sub-regions of the input space represent the heterogeneity. The disconnected sub-regions that can be characterized by the same function are said to be in the same class, and evaluating the function at input returns the minimum distance to a boundary of the contiguous class (sub-region).
The ClassGP modeling framework, previously developed to model for such heterogeneous functions along with a novel MCMC based tree sampling algorithm, is used to introduce a novel tree-based optimization framework dubbed as ClassBO (Class Bayesian Optimization). We present theoretical analysis of the algorithm and demonstrate the superior performance of ClassBO against other methods via empirical evaluations

3 - A Parametric Approach for Solving Convex Quadratic Optimization with Indicators Over Trees

Aaresh Bhathena, University of Michigan, Ann Arbor, MI, United States, Salar Fattahi, Andres Gomez, Simge Kucukyavuz

This paper investigates convex quadratic optimization problems involving n indicator variables, each associated with a continuous variable, particularly focusing on scenarios where the matrix Q defining the quadratic term is positive definite and its sparsity pattern corresponds to the adjacency matrix of a tree graph. We introduce a graph-based dynamic programming algorithm that solves this problem in time and memory complexity of $\mathcal{O}(n^2)$. Central to our algorithm is a precise parametric characterization of the cost function across various nodes of the graph corresponding to distinct variables. Our computational experiments conducted on both synthetic and real-world datasets demonstrate the superior performance of our proposed algorithm compared to existing algorithms and state-of-the-art mixed-integer optimization solvers. An important application of our algorithm is in the real-time inference of Gaussian hidden Markov models from data affected by outlier noise. Using a real on-body accelerometer dataset, we solve instances of this problem with over 30,000 variables in under a minute, and its online variant within milliseconds on a standard computer. A Python implementation of our algorithm is available at (https://github.com/aareshfb/Tree-Parametric-Algorithm.git).

4 - Global Multi-Objective Simulation Optimization with Low Dispersion Point Sets

Burla Ondes, Purdue University, West Lafayette, IN, United States, Susan Hunter

Consider the context of a Multi-Objective Simulation Optimization (MOSO) problem, that is, an optimization problem with multiple simultaneous and conflicting objectives that can only be observed with stochastic error through a Monte Carlo simulation oracle. The goal of solving a MOSO problem is to identify decision points that map to the global Pareto set. For such problems featuring continuous objectives on a compact feasible set, we determine how one should trade off the number of decision points to sample, m, with the number of simulation replications per sampled point, n, to ensure the estimated Pareto set converges to the true Pareto set under a novel performance measure called the modified coverage error. This performance measure enables us to quantify upper bounds on the deterministic and stochastic error as a function of the dispersion of the m points in the decision space and the number of simulation replications per point, n. The upper bounds and tradeoff analysis lead to efficient algorithms for solving global MOSO problems with probabilistic guarantees using low dispersion point sets.

5 - Optimizing Camera Placement in Critical Environments: A Simulation Optimization Approach

Yash Kumar, The University of Texas at Austin, Austin, TX, United States, Raghu Bollapragada, Benjamin Leibowicz

Camera placement optimization is critical for surveillance, environmental monitoring, and infrastructure inspection, yet even simplified versions are NP-hard due to high-dimensional camera configurations and challenges in accurate coverage calculation. This paper presents a novel simulation optimization framework for camera placement, combining discrete and continuous optimization methods for efficient free space coverage. Our approach employs mixed-integer programming and derivative-free optimization in an iterative structure, segmenting the free space into weighted voxels (3D pixels). By refining camera configurations based on visibility approximation and adaptive sampling, our method maximizes visibility coverage while considering resource constraints. Fine-tuning with a granular spatial grid improves accuracy, and utilizing derivative-free optimization methods improves the objective even further. Our research emphasizes intelligent strategies to enhance camera placement efficiency and robustness. Through case studies, we demonstrate the effectiveness of our approach in critical environments where efficient surveillance is crucial.

WB33

Summit - 424

Stochastic Integer Programming with Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Juan-Alberto Estrada-Garcia, University of Michigan, Ann Arbor, MI, 48109, United States

1 - Distributionally Robust Multimodal Generation Maintenance and Operations Scheduling with Storage and Wind Energy Uncertainty

Natalie Randall, University of Iowa, Iowa City, IA, United States, Beste Basciftci

This study focuses on a predictive generator maintenance and operations scheduling problem with storage units under multimodal wind energy uncertainty. We formulate this problem as a two-stage distributionally robust mixed-integer program under a multimodal moment-based ambiguity set, where the first-stage problem determines the maintenance and unit commitment related decisions of the generation units, whereas the second-stage problem determines the production related plans of the generators and storage units under the ambiguity of wind energy. We present a column-and-constraint generation algorithm to solve our large-scale problem. We provide a computational study over different IEEE bus instances, and analyze the impact of wind farms and storage units on the system while demonstrating the computational efficiency of the solution algorithm.

2 - Neural Approximate Dynamic Programming for the Ultra-Fast Order Dispatching Problem

Mucahit Cevik, Toronto Metropolitan University, Toronto, ON, Canada

Same-Day Delivery (SDD) services aim to optimize the speed of online order fulfillment while taking into account operational challenges such as fluctuations in order volumes and uncertainties in courier schedules. Our work enhances the effectiveness of SDD by concentrating on the ultra-fast Order Dispatching Problem (ODP), which entails the rapid assignment and dispatching of orders to couriers within a central warehouse environment, aiming to complete deliveries within tight deadlines (e.g., within minutes). We introduce improvements such as order grouping and specific courier allocations, offering a more accurate depiction of dispatch operations and boosting delivery performance. Our approach primarily employs NeurADP, a novel technique that merges Approximate Dynamic Programming with Deep Reinforcement Learning (DRL), marking its first application beyond ride-pool matching. NeurADP effectively handles the complex matching and routing challenges inherent in ultra-fast ODP through a neural network-based Value Function Approximation that adeptly manages high-dimensional problem complexities without the need for traditional manual feature engineering. We evaluate our method on four specialized datasets designed for ODP, assessing NeurADP's performance against standard myopic and DRL benchmarks and utilizing derived bounds to examine policy effectiveness. The numerical results show that incorporating order batching and courier queues significantly boosts delivery efficiency, with NeurADP outperforming alternative approaches.

3 - Stochastic Integer Programming with Decision-Dependent Uncertainty for Infrastructure Monitoring

Siqian Shen, U of Michigan/NSF, Ann Arbor/Alexandria, MI, United States, Juan-Alberto Estrada-Garcia

Infrastructure monitoring is critical for systems that are subject to uncertain failures. We consider the problem of dispatching multiple maintenance fleets (e.g., drones) to examine different parts of a system, whose failures are probabilistic and also depend on inspection schedules and fleets' arrival time. The goal is to balance the fleets' routing and operational cost with the uncertain cost of system failure and maintenance. We model a two-stage stochastic mixed-integer program with decision-dependent uncertainty and solve it via a stochastic decomposition algorithm. We speed up the computation of subproblems using column generation and random coloring algorithms. Furthermore, we consider dynamic system monitoring and real-time drone routing with updated system information after each arrival, and solve the problem using stochastic dual dynamic integer programming and reinforcement learning. We conduct numerical studies using networks with different sizes and topologies, as well as the IEEE 33-bus system, to demonstrate the efficiency and effectiveness of different approaches.

4 - Multi-stage stochastic programming models for long-term disaster housing planning

Sheng-Yin Chen, Clemson University, Central, SC, United States, Yongjia Song

In the long-term disaster housing logistics planning, our objective is to establish a housing logistics network for sustained disaster response and provide a long-term inventory strategic policy framework that decision-makers. We apply various types of multi-stage stochastic programming models to capture the life-cycle of disaster housing and aim to balance between long-term logistics cost and the quality-of-life of disaster victims.

WB34

Summit - 425

Advanced Scheduling and Optimization Techniques in Value Chains

Contributed Session

Chair: Payman Jula, Simon Fraser University, Vancouver, BC, Canada

1 - Scheduling parallel non-homogeneous machines in coordinated value chains.

Payman Jula, Simon Fraser University, Vancouver, BC, Canada

Motivated by scheduling challenges in complex manufacturing environments, we propose mathematical programming models and heuristic solutions to schedule parallel non-homogeneous machines that must be coordinated with other areas of the chain. We address selecting and scheduling jobs considering sequence-dependent setup times and strictly enforced time windows. The efficiency, effectiveness and robustness of the proposed methods are discussed.

2 - A Mixed-Integer Optimization Tool for System-Level Proactive Obsolescence Risk Management

Chad Uhles, University of Tennessee, Knoxville, Knoxville, TN, United States, Hugh Medal

Diminishing manufacturing sources and material shortages (DMSMS) pose a significant risk to the reliability and maintainability of many modern complex systems due to part obsolescence. Part obsolescence increases the risk of system failure or downtime, which is often not only undesirable, but non-negotiable. As such, it is imperative to anticipate events of part obsolescence and take proper actions to mitigate their effects. However, developing a detailed system-wide plan of mitigating events of obsolescence could necessitate the consideration of thousands, or even millions, of decisions before an optimal plan is derived. Thus, there exists a need for an automated tool that can produce a cost optimal obsolescence management plan. In this work, we propose a mixed-integer optimization tool that proactively manages and mitigates events of obsolescence at the system level. By considering the architecture of the bill of materials, part inventories, expected system demand, and forecasted obsolescence dates, our tool makes decisions of when and how to enact up to 8 unique short-term and long-term resolutions for every part within the system at every time until the anticipated system end of life. The parameters determine the structure of a mixed-integer optimization problem, where the objective is to minimize costs for the system over the entire time horizon. We present the findings of a case study that compares the obsolescence management plan of a reactive strategy to a plan produced by our mixed-integer programming optimization tool and derive insights based on the difference between the solutions produced by the two methods.

3 - Cooperation and competition in two-agent scheduling with parallel batching

Renjie Yu, The University of Sydney, Camperdown, Australia, Daniel Oron

We study single-machine scheduling problems involving two competing agents sharing a batching machine. This machine can handle at most one job from each agent simultaneously. Our focus is on two due date-related scheduling criteria: the number of late jobs and the total late work. We investigate two variations for each problem: minimizing the combined objectives of both agents, and minimizing the objective of one agent while imposing an upper bound on the other agent's objective. All four versions of the problem are demonstrated to be strongly NP-hard. To address this complexity, we devise efficient Mixed-Integer Linear Programming (MILP) formulations, and employ constructive algorithms and Tabu Search for approximate solutions. A comprehensive numerical analysis validates the effectiveness of the approximation algorithms.

4 - A restless bandit approach for capacitated condition-based maintenance scheduling

Ece Zeliha Demirci, TED University, Ankara, Turkey, Joachim Arts, Geert-Jan Van Houtum

We address the maintenance scheduling problem of multiple non-identical machines deteriorating over time. As the machines deteriorate, their efficiency diminishes, leading to reduced output quality and revenue losses. Maintenance costs increase as the machines' condition worsens, and the number of simultaneous maintenance tasks is restricted by the number of maintenance workers. We cast the problem as a restless bandit problem and propose an efficient index-based heuristic, namely Whittle’s index policy. Additionally, we establish a lower bound on the optimal solution using linear programming. Our numerical experiments demonstrate that Whittle’s policy outperforms both failure-based and threshold-based maintenance strategies.

WB35

Summit - 427

Platforms Operation in Urban Transportation

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Sisi Jian, HKUST, Clear Water Bay, N/A

1 - Stackelberg Game in Mobility-as-a-Service Platforms with Stochastic Matching between Users and Operators

Bingqing Liu, New York University, Brooklyn, NY, United States, David Watling, Joseph Chow

This study proposes a stochastic Stackelberg game to model Mobility-as-a-Service (MaaS) systems, where the platform is the leader, users and operators are the followers. The platform set fares to maximize revenue. Users and operator react to the fares to form a stochastic many-to-many assignment game considering both fixed-route services and Mobility-on-Demand (MOD). There are two major contributions. First, the many-to-many assignment game is extended into a stochastic form while including the impact of fares. The stochastic assignment game captures the heterogeneity in both users and operators with a coalitional route-choice logit model. Second, the role of the platform is modeled as the leader of the platform who does pricing while fully anticipating the reaction of users and operators with the Stackelberg game. The Stackelberg game is written into a bilevel problem. The lower level is the stochastic many-to-many assignment game between users and operators, which is shown to yield a coalitional logit model. The upper level sets fares to maximize revenue. An iterative balancing algorithm is proposed to solve the lower-level problem. The bilevel problem is solved through an iterative fare adjusting heuristic with a Frank-Wolfe fare adjusting mechanism, whose solution is shown to be equivalent to the bilevel problem with an additional condition when it converges. Two case studies are conducted. The model could be applied to design MaaS fares maximizing income of the platform while anticipating the selfish behavior and heterogeneity of users and operators. Public agencies could also use the model to manage multimodal transportation systems.

2 - Preference-based matching mechanism tailored for ride-hailing platforms considering congestion effects

Yanyan Ding, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Shixuan Tang, Sisi Jian

With the widespread and rapid growth of on-demand e-hailing service, how to effectively match passengers and drivers becomes one of the most significant operational problems for online platforms to realize their best revenues. This study therefore investigates the e-hailing matching problem from the perspective of platforms. Considering the passenger preference, a menu-based assignment mathematical model is formulated to maximize the expected operational revenue. For tractability, a deferred algorithm (DA) to effectively assign large-scale instances. Numerical experiments are conducted based on real data from Haikou, Hainan to validate the effectiveness and efficiency of the proposed method, followed by some practical insights and implication for the online matching platforms.

3 - The VCG Pricing Policy with Unit Reserve Prices for Ride-Sourcing is 3/4-Individual Compatibility

Xiaotong Sun, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Hong Kong, Ruijie Li, Haiyuan chen, Kenan Zhang, Xiaobo Liu

We consider a customized pricing and matching strategy for ride-sourcing services in which drivers and riders disclose their willingness to pay (WTP) per unit distance and willingness to accept (WTA) per unit distance, respectively. The platform then matches drivers and riders based on their reports to maximize the total social welfare. While this strategy is motivated to promote transparency, fair pricing, and participant satisfaction, it faces a particular challenge to ensure riders and drivers truthfully report their valuations, which are considered private information. To encourage truthful reports, we implement the Vickrey-Clarke-Groves pricing policy augmented with unit reserve prices, referred to as VCG-URP. For the riders, the reserve price is set as the minimum WTP per unit distance, while for the drivers, it is the maximum WTA per unit distance. We demonstrate that VCG-URP ensures that riders provide truthful valuations. However, it does not guarantee the same level of honesty from drivers in reporting their valuations. While drivers cannot enhance their utility by overstating their unit WTA, they may benefit from understating it. These theoretical results are confirmed by numerical experiments with several additional findings: i) although the underreporting strategy of the drivers theoretically has an opportunity to achieve a higher utility, it always leads to a decrease in expected utility over randomly generated matching scenarios; ii) the introduction of reserve prices has larger impacts on riders than drivers and benefits the platform in excluding less profitable trips.

4 - eMaaS design that incorporates the three-sided matching among traveler, operator, and energy provider

Hai Yang, New York University, Brooklyn, NY, United States, Bingqing Liu, Joseph Chow

In the electric Mobility-as-a-Service (eMaaS) market, travelers interact with mobility providers, and the latter in turn interact with energy providers. This study proposes an electric-MaaS (eMaaS) assignment game capturing interdependencies between operators, travelers, and energy providers. The problem is formulated into a many-to-many 3-sided assignment game, using the supplier-firm-buyer structure. Matching is represented by two sets of disjoint path flows: flows of travelers along paths for ODs, and flows of vehicles to energy providers. The model captures steady state probabilistic flows. The matching problem can be represented by a tractable combination of a path-based multicommodity network design problem and a transportation problem. To capture probabilistic charging demand, we adopt a stochastic assignment game via a coalitional logit model where disjoint coalitions (travelers + mobility operators, mobility operators + energy providers) choose among a set of paths using a logit model. We propose to model the stability constraints corresponding to this problem as optimality conditions of the stochastic user equilibrium (SUE) problem, and the stochastic assignment game transforms into a Stackelberg game with platform pricing leaders (energy prices, mobility prices) and matching mobility operators, energy providers, and travelers as followers. The upper-level problem determines the pricing of service fares and charging costs, while the lower-level problem solves the assignment game between operators and travelers, as well as the assignment between operators and energy providers. By changing the upper-level objective function (max consumer surplus vs. max system profit), different pricing strategies of both operators and energy providers can be assessed.

WB36

Summit - 428

Modeling and Optimization in Transportation and Mobility Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Wang Chen, The University of Hong Kong, Hong Kong, N/A

1 - Joint Pricing, Matching, and Repositioning for Dynamic High-Capacity Ride-Sharing Considering Future Opportunity Costs

Wang Chen, Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, China, People's Republic of, Jintao Ke

This research proposes a joint optimization framework for dynamic high-capacity ride-sharing, which simultaneously optimizes pricing, matching, and repositioning considering future opportunity costs that indicate the future expected revenue. This research considers passengers to be price-sensitive and have a choice to decline the upfront prices offered by the platform. The future opportunity costs and passengers' choice behavior are integrated into the optimization framework that maximizes the platform's expected revenue with given batches of passengers and drivers. This research models the framework as a two-stage stochastic optimization problem. This problem is relaxed to a one-stage problem with a convex objective function, achieving comparative accuracy while dramatically lower computational costs. Numerical studies show that the proposed optimization framework can create substantial economic benefits for the platform in different cities without adjusting hyperparameters, indicating the proposed optimization framework could be generally applied to other cities.

2 - Competition in the Intercity Passenger Transport Market with Ride-Pooling Services

Jinhua Si, Department of Industrial Engineering, Tsinghua University, beijing, China, People's Republic of

Ride-pooling services are increasingly promising for intercity travelers. These demand-responsive services are notably effective in promptly addressing intercity requests and reducing overall travel durations. However, planning for this emerging mode must also account for heightened competition from existing modes, such as conventional intercity buses and the expanding high-speed rail networks. In this study, we develop stylized game models to characterize the intercity passenger transportation market, which includes ride-pooling services and alternative intercity transportation modes. Through hierarchical passenger segmentation based on travel urgency and analysis of demand preferences across urban areas, we optimize fleet size and trip fares to achieve market equilibrium. By assessing competitive advantages across various market parameters and spatial conditions, we highlight the transformative impact of ride-pooling services on traditional modes and propose strategic entry points to leverage latecomer advantages. Our findings have implications for the deployment of intercity demand-responsive services, aiming to enhance benefits for both service providers and intercity travelers.

3 - Railway Disruption Management: Designing Shuttle Bus Services and Passenger Assignment with Subsidized Taxi Support

Hongzheng SHI, Department of Civil Engineering, The University of Hong Kong, Hong Kong, China, People's Republic of, jiguang Wang, Jintao Ke

Rail transit systems are playing an increasingly important role in major cities worldwide, leading to more frequent service disruptions. Shuttle bus service is a common coping strategy for these disruptions, but it suffers less flexible, which may lead to higher operating costs to serve all affected passengers. Giving subsidies to taxi drivers near the disruption location can attract them to provide flexible capacity for the shuttle bus service and thereforme reduce the total service cost while meeting passenger's demand. In this paper, we systematically investigate the problem of designing shuttle bus service with subsidized taxi support, which involves determining the taxi subsidy amount, shuttle bus route design, frequency determination, and passenger assignment. The problem is formulated as a mixed-integer programming (MIP) model with the objective of simultaneously minimizing route operating time, detour penalty, shuttle bus usage cost, and taxi subsidy cost. To address the model's nonlinearity, we propose a predict-then-optimize framework that first predicts the optimal subsidy and then optimizes the shuttle bus service design with the predicted subsidy. A column-generation-based algorithm is proposed to efficiently solve the model after the prediction process. Experiments demonstrate that our method can provide efficient response plans in advance for various disruptions.

4 - A Two-Stage Stochastic Programming Optimization for Multi-Type Chargers Deployment and Electricity Distribution

jiguang Wang, Department of Civil Engineering, the University of Hong Kong, Hong Kong, China, People's Republic of, Hongzheng SHI, Xinglu Liu, Kefan Lai, Jintao Ke

Electric vehicle systems have emerged as a promising and sustainable solution for urban mobility. This study proposes an integrated optimization problem that addresses the long-term deployment of multi-type charging facilities for electric taxi systems at the planning level. Additionally, it optimizes electricity distribution decisions from energy storage systems (ESS) and the power grid at the operational level, while considering the nonlinear operating costs of charging power supply. Charging demand is estimated from real taxi trajectory data, then to account for demand uncertainty, a two-stage stochastic integer program is formulated, and valid inequalities are designed to expedite the exact solution of the model. To handle a large number of scenarios in the stochastic program, the sample average approximation scheme is employed as the sampling strategy. The modeling and algorithm approach is rigorously tested through a comprehensive case study conducted in Hong Kong, China. The results of extensive experiments demonstrate that the deployment of multi-type charging infrastructure, along with the joint optimization of ESS and the power grid, can significantly enhance system profitability and improve operational performance. Sensitivity analysis is utilized to provide valuable management recommendations.

WB37

Summit - 429

Advances in Delivery Service Operations

Contributed Session

Chair: Mouna Bamoumen, Tilburg University, Tilburg, N/A, Netherlands

1 - The Location-Routing Problem with Pickup Facilities and Heterogeneous Demand: Formulation and Heuristic Approach

Yida Xu, Tianjin University, Tianjin, China, People's Republic of, ZHAO ZHAO, Zhaofang Mao, Kan Fang

Nowadays, as more uncertain influential factors have emerged and customer demand becomes more heterogeneous and discrete, last-mile delivery has become a time-consuming and uncertain process. Unreasonable distribution network design would lead to higher delivery failure rates and lower vehicle utilization. As a result, courier companies have delpyed various innovative parcel storage facilities, including pick-up points and lockers.

In order to optimize the last-mile distribution network to adapt to the complex and ever-changing distribution environment, we focus on the introduction of self-pickup facilities in modern logistics and distribution scenarios under heterogeneous customer demands. We distinguish the demand into two categories: general goods and special goods (e.g., irregular package and refrigerated goods), as well as corresponding transportation vehicles. Subsequently, we address the location-routing problem with pickup facilities and heterogeneous demands (LRP-PFHD). To solve this challenging problem, we propose a mixed integer linear programming (MILP) model that aims to minimize the total cost. Due to the NP-hardness of the problem, we propose an adaptive large neighbourhood decomposition search algorithm to solve LRP-PFHD. We evaluate both the effectiveness and the efficiency of the proposed method using generated instances based on benchmark datasets. The results demonstrate the effectiveness of the proposed method compared to commercial MILP solvers such as Gurobi for LRP-PFHD, especially for large-scale instances. In addition, we make a comprehensive analysis of certain parameters (e.g., facility opening cost, transportation cost) to explore their impacts on the results and provide helpful managerial insights for courier companies.

2 - Combining Gamification and Crowdsourcing to Improve Solutions for the Container Loading Problem Applied to Large Objects

Jorge Elias Garcia Zamalloa, Georgia Institute of Technology, Atlanta, GA, United States, julien Maurice, Benoit Montreuil, Ali Barenji

The last-mile delivery problem is one of the key hot topics that researchers around the world are studying. Its relevance is mainly triggered by the costs associated with it, the rapid increase of e-commerce, and customers who expect to receive their orders almost instantly. Under this setting, an alternative that is becoming more popular is the introduction of modular warehouses—semi-permanent engineered infrastructures that can be assembled in a few weeks and can be disassembled to install elsewhere. Additional benefits include reduction of setup costs and closing the distance between companies and the market.

However, the main drawbacks are the difficulty of transporting the structure, and the capacity of the trucks to carry it efficiently. An optimization approach can be used to generate the required loading plan, but it is a resource-consuming process since it is a combinatorial optimization problem. Moreover, since most objects involved tend to have irregular shapes, the heuristics needed to solve the problem usually incorporate simplified geometric constraints and tend to generate suboptimal solutions.

We propose to generate an initial non-optimal containerization plan and improve it by leveraging human factor, gamification, and crowdsourcing. The key idea is to create ludic and realistic scenarios where people perform the containerization of 3D irregularly shaped objects and compete to maximize the utilization of the vessel. Then, an agent can learn from the best proposed configurations and develop its own strategies for new settings. This approach creates new research directions for the container loading problem.

3 - A Hybrid Q-Learning Evolutionary Approach for Routing Planning in Mobile Store Services

Chieh-Ju Juan, Department of Aviation Management, Air Force Academy, Taiwan, Kaohsiung, Taiwan, Taiwan

In recent years, with the rapid development of technological advantages and the concept of sharing economy, the prevalence of online delivery platforms has surged, providing people with convenient and fast options for fulfilling their consumption preferences. In order to bolster the competitiveness in sales, traditional retailers have begun introducing mobile store services. These mobile vehicles utilize their flexibility to park at different locations, thereby leveraging their adaptability to cater to a wide range of areas, effectively meeting the needs of a larger customer demographic. In this study, we define the Mobile Store Vehicle Routing Problem (MSVRP) as a dynamic vehicle routing problem with time windows. The objective is to enhance the adaptability of routing planning for mobile store vehicles, enabling them to effectively navigate the increased complexity resulting from fluctuating environmental conditions. A hybrid Q-Learning evolutionary algorithm (HQEA) is developed to solve the MSVRP, where two objectives are simultaneously considered: total duration of vehicles, and operation profit. In HQEA, the nondominated sorting genetic algorithm II serves as the static optimizer in stable environments, while a continuous reoptimization procedure is devised to address dynamic conditions. Within this reoptimization framework, three strategies are introduced and incorporated into a Q-learning mechanism, functioning as an agent to guide transitions between strategies. This approach aims to enhance the algorithm's performance systematically. The experimental results obtained from testing on benchmark datasets demonstrate that the proposed algorithm is successful in generating promising routing plans.

4 - Dynamic pricing and routing for online food delivery

Mouna Bamoumen, Tilburg University, Tilburg, Netherlands, Haitao Liu, Jan Fransoo

We examine the case of dynamic pricing and routing for online food delivery, where a group of couriers transports goods from various restaurants to customers who place orders. This work is motivated from a direct observation on the potential delays caused by prolonged waiting times for elevators during peak hours. To incorporate this new aspect into the modelling, our objective is to decide on delivery alternatives' prices and to allocate orders to couriers in order to efficiently manage the fleet of drivers and meet customer deadlines with the least amount of delay. This problem is characterized by three main sources of uncertainty. First, customers' identities remain unknown until they initiate an order. Additionally, the choice of delivery option is not known until the customer confirms their order. Lastly, the service time is impacted by the waiting time for elevators, which is not observed until the courier reaches their destination. We model this integrated demand management with dynamic pricing and routing decision problem under stochastic service times for Same-Day Delivery as a Markov Decision Process. We propose a Value Approximation Function to find the optimal policy of pricing and routing that maximizes the total expected revenue.

WB38

Summit - 430

AI-Driven Healthcare and Public Engagement

Contributed Session

Chair: YIXUAN LIU, Hong Kong Polytechnic University, hongkong, N/A, China, People's Republic of

1 - Auto-Development of AI Systems for Predicting Discard of Donated Organs Empowered by Deep Reinforcement Learning

Yuzhang Han, California State University San Marcos, San Diego, CA, United States, Zahra Gharibi, Mehmet Ayvaci

We develop an advanced Artificial Intelligence (AI) system to detect procured organs at a heightened risk of being discarded. The AI identifies these organs and investigates the underlying reasons for potential discards, categorizing them into five distinct groups based on donor clinical, behavioral, and social characteristics. This categorization provides critical insights into the primary factors contributing to organ discards, enabling the development of improved strategies for organ handling and allocation. The AI model aids decision-makers and policy designers in enhancing organ utilization and patient outcomes by facilitating strategic organ allocation to efficiently match viable organs with recipients, thus reducing unnecessary discards and improving the transplantation process.

2 - Medical Livestreaming as Pipes and Prisms on Online Medical Consultation Platforms

YIXUAN LIU, Hong Kong Polytechnic University, hongkong, China, People's Republic of, Xiaofei Zhang, Mike Lai

Medical live streaming, a utilization of live streaming technology in online medical consultation (OMC) platforms, can be a bridge to close the information gap and alleviate medical disparities. Many OMC platforms have introduced this feature allowing qualified physicians to conduct medical live streaming. However, the impact of medical live streaming on platform operations remains unclear. Specifically, medical live streaming may present a paradox: it may help clarify uncertainties, and encourage patient consultations and patient professionalism; or it could promote self-diagnosis, reducing the patients' desire to consult and reduce patient professionalism. From a theoretical standpoint, there is also limited understanding regarding how the role of medical live streaming diverges from that of conventional live streaming. Therefore, it is crucial to investigate how medical live streaming influences patient consultations and patient professionalism, as well as the underlying mechanisms. In this paper, we study these tradeoffs through a quasi-natural experimental setting. We show that medical live streaming has a significant, positive, and robust impact on platform operations, in terms of patient consultations and patient professionalism. We further explore the underlying mechanisms and find that the impact is exerted potentially match-up effect. We also conduct a series of analyses to unravel the heterogeneity associated with the characteristics of live streaming and physicians/patients. Our findings contribute to the understanding of the role of medical live streaming and provide insights to academics and practitioners.

3 - Towards a robust decision aid for personalized deceased-donor kidney offer evaluation

Fatemeh Gholizadeh, University at Buffalo, Buffalo, NY, United States, Prashant Sankaran, Robert Dell, Moises Sudit, Liise Kayler

Patients with end-stage kidney disease (ESKD) primarily have two options for survival: dialysis and kidney transplant. When patients with ESKD opt for a kidney transplant, they join the dynamic national transplant waiting list where they are assigned an allocation score based on their characteristics that updates with health changes. When a deceased donor organ becomes available for transplant, its viability is assessed, a matching algorithm is run to find compatible candidates from the waiting list, and the kidney offer is made to the highest-ranked candidate. This offer includes information about the kidney to help the patient and their physician in evaluation. If rejected, the offer proceeds to the next candidate. While assessing the kidney offer, the physician must inform the candidate about the tradeoff in accepting or rejecting the offer. Estimating the quality and waiting time of a future kidney offer are two critical pieces of the decision-making process. Physicians generally cannot provide quantitative estimates of expected wait times and the quality of future kidney offers. Further, the offers available to lower-ranked candidates are affected by the decisions made by the higher-ranked candidates, highlighting the interdependence among the patients on the waiting list. The existing literature on personalized estimation of waiting times and the quality of future offers has not considered the influence of ranking on waiting times. Our work addresses this gap by including estimates of a candidate's ranking to estimate their waiting time and the quality of future offers.

WB39

Summit - 431

Advancements in Medical Technology and Healthcare

Contributed Session

Chair: Carolina Arroyo-Roldan, University of Puerto Rico - Mayaguez, Añasco, PR, 00610, United States

1 - Feature Subset Selection Framework for Endometriosis Diagnostics

Carolina Arroyo-Roldan, University of Puerto Rico - Mayaguez, Mayaguez, PR, United States, Wandaliz Torres-Garcia

Endometriosis is a gynecological disease characterized by the growth of endometrial tissue outside of the uterus. It is known to cause pelvic pain, dyspareunia, and dysmenorrhea, and has been linked to infertility. The diagnosis of endometriosis takes various years, partly because the gold standard for diagnosis is an invasive laparoscopic surgery which poses both a physical and financial burden on women. Previous research studies to identify potential molecular biomarkers face statistical and computational limitations due to small sample sizes with high dimensionality. Additionally, there is no standardization regarding factors that should be considered when analyzing the gene expression of endometrial tissue, such as disease severity and the phase of the menstrual cycle. Though experiments with large sample sizes are not available, many new and upcoming experiments are being generated to bridge the gap in understanding the biological mechanisms of this largely uncharacterized disease affecting many women worldwide. Therefore, we propose the development of a consensus framework to extract the best feature subsets to discriminate between endometriosis patients versus women without the disease using gene expression of endometrial tissue, as well as clinical information such as the stage of the menstrual cycle. This framework aims to identify and rank feature subsets using (1) optimization-based search algorithms coupled with classification methods and (2) queries of functional processes from gene-annotated databases across multiple publicly available experiments from the Gene Expression Omnibus repository to evaluate consensus.

2 - Deep Learning to Synthesize Dynamic Contrast-Enhanced Breast MRI for Breast Cancer Screening

Suleeporn Sujichantararat, Department of Radiology, University of Washington, Seattle, WA, United States, Debosmita Biswas, Anum Kazerouni, Michael Hirano, Habib Rahbar, Savannah Partridge

Breast magnetic resonance imaging (MRI) is the most sensitive tool for breast cancer detection. Conventional breast MRI relies on gadolinium-based contrast agents (GBCAs) to produce dynamic contrast-enhanced (DCE) images, which reflect abnormal tissue vascularity and aid physicians in identifying cancers. However, GBCAs may pose side effects to patients and repeated doses used in annual breast MRI screening pose unknown health risks. Furthermore, administration of GBCAs is costly and time consuming, which makes MRI exams overall less accessible for general breast screening. To eliminate the need for GBCAs and improve the safety and accessibility of breast MRI, prior work has successfully utilized Artificial Intelligence (AI), specifically deep learning models, to synthesize DCE images from unenhanced images. In this work, we aim to refine the existing models, which were primarily designed for demonstrating larger invasive breast cancers, to detect early cancers in a screening population of women with dense breasts. One approach is utilizing U-Net, a convolutional neural network developed for biomedical image segmentation. Another approach is utilizing a hierarchical fusion network with attention mechanism model. Inputs to these models include a variety of combinations of different unenhanced MR sequences: T1-weighted, T2-weighted, diffusion-weighted imaging (DWI) with different b-values, and apparent diffusion coefficient (ADC) map. We will train and evaluate models on a dataset of women with mammographically dense breast tissue undergoing MRI screening. Evaluation metrics will include quantitative assessment using Structural Similarity Index Measurement (SSIM) and Normalized Mean Squared Error (NMSE), and qualitative multireader radiologist assessment of synthetic DCE quality.

3 - Combating the Societal Grand Challenge of Ransomware: Impacts of Disclosed Organizational Responses from an Epidemiological Perspective

Shivan Bhatt, Clemson University, Clemson, SC, United States, He Li, William Kettinger

Ransomware is a form of malware that involves three, sequential events. Hackers access a system, attack the system by preventing access and/or encrypting data, and demand a ransom in return for providing access to the system or decrypting data i.e., digital extortion. Ransomware attacks inflict significant costs on organizations, such as financial, downtime, opportunity, and reputation costs, while hackers use the ransom to fund more attacks, underscoring the importance of understanding ransomware and developing managerial guidance. Ransomware exists as strains, indicating that the malware contains a shared codebase with other malware. Survival of a ransomware strain is influenced by how organizations respond to ransomware attacks e.g., paying a ransom or coordinating with security experts. Organizational responses to ransomware attacks affect hackers' decision-making calculus i.e., how hackers perceive costs and benefits of strains. Hackers select strains that offer the highest benefit and lowest cost. We examine organizational responses to ransomware attacks and their impact on strain survival. Using time-to-event analysis, we find that both the immediate and cumulative effect of paying a ransom is an increase in strain survival. This contrasts with the effect of coordination, defined as an organization either engaging with law enforcement or information security experts. The immediate effect of coordination does not impact strain survival; only the cumulative effect of coordination decreases strain survival. We also find that responses are not path dependent i.e., previous responses do not predict current responses. These findings provide important theoretical and practical insight for organizations responding to ransomware attacks.

4 - Patient Satisfaction in Healthcare: The Roles of Pre-defined Service Aspects and Unstructured Comments

Lin Lu, Fairfield University, Fairfield, CT, United States, Qiong Hu

Healthcare services have undergone transformative changes to adapt to new safety protocols and technological advancements in response to the COVID-19 pandemic. This study investigates whether pre-defined service dimensions continue to capture the full spectrum of patient experiences or if new, emergent needs are being highlighted by patients. By analyzing data from medical surveys that include both established service dimensions and unstructured comments, we aim to identify shifting paradigms in patient expectations and ensure continuity of care. Our findings will inform potential improvements in service design, enhancing patient safety and satisfaction in the post-pandemic era.

5 - Predicting Sepsis Mortality using Machine Learning

Ameera Ibrahim, Saint Mary's College of California, Moraga, CA, United States, Ahmed Ahmadein

Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated host response to infection. According to the World Health Organization (WHO), it is estimated that 27% of patients with sepsis in hospitals and 42% of sepsis patients in intensive care units will die. With such a high mortality rate, sepsis has gained its reputation as an important public health issue and motivated many researchers to work on improving the prevention, diagnosis and treatment of sepsis. In this work, we propose a novel methodology to predict mortality in patients with sepsis. Our main goal is to design a reliable tool that can assist clinicians in evaluating the probability of a patient's fatality and identifying the best treatment plan for patients with sepsis. We exploit the XGBoost algorithm, a leading machine learning technique, on data extracted from the MIMIC-III dataset, a large and freely-available database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units.

6 - Modeling Single-Donor Organ Chain Exchange: Optimization Framework for Analyzing Transplant Success

Yunxia (Peter) Zhu, University of Nebraska-Lincoln, Lincoln, NE, United States, Rakesh Mallipeddi, Chelliah Sriskandarajah, Jon Stauffer

An optimal matching problem to maximize transplant success rate is investigated in the context of single-donor organ exchange, where one living donor's organ is transplanted to a single patient. In such exchanges, multiple medical compatibility criteria pose a serious challenge for matching a patient with compatible donor. The objective of our study is to increase matching donors with patients requiring organs with maximum transplant success rate. In order to facilitate this organ matching, we use chain mechanism to increase the number of patient-donor matches with a maximum success rate for patients receiving transplants. We formulate a general optimization framework for obtaining the chain that maximizes the success rate of patients receiving transplants taking into account multiple compatibility criteria.

WB40

Summit - 432

Innovations in Healthcare Service Delivery

Contributed Session

Chair: ILGIN ACAR, Western Michigan University, Department of Industrial and Entrepreneurial Engineering and Engineering Managem, Kalamazoo, MI, 49008-5336, United States

1 - Optimal contract policies for improving access to child and youth public services: the case of the Ontario Autism Program

Saha Malaki, Western Ontario University, London, ON, Canada, Felipe Rodrigues, Salar Ghamat, Camila de Souza, Greg Zaric

Access to public services in Canada has declined significantly, with the Ontario Autism Program (OAP) as a prime example. This publicly funded program currently has more than 60,000 children registered and waiting up to two years for provincial funding to access decentralized private services and treatments for autism spectrum disorder (ASD). We focus on the strategic decisions that families, service providers, and governments make in a decentralized system characterized by conges:on, where competing interests have been introduced in asymmetric competition, providing a strategic queuing game behaviour among its players. We developed a general mathematical framework to explore the structure of optimal contractual agreements for the OAP in the presence of congestion and information asymmetry in offering public treatment services for autism. This preliminary framework can then be generalized to other decentralized congested healthcare services.

2 - Last-Mile Attended Home Healthcare Delivery: a Robust Strategy to Mitigate Cascading Delays and Ensure Punctual Services

Mingda Liu, Tsinghua University, Beijing, China, People's Republic of, Yanlu Zhao, Xiaolei Xie

The attended home healthcare (AHH) industry is witnessing rapid growth due to the rising demand from an aging population and the potential benefits of alleviating pressures on traditional healthcare resources. However, ensuring timely one-on-one AHH services for homebound patients remains a challenge because of cascading delays arising from uncertainties in travel and service times. To address this issue in the last-mile homecare delivery, we develop a systematic cascading delay mitigation strategy that ensures patients receive dependable homecare services. Specifically, we introduce a compound set reliability index (CSRI) that captures risk exposure by characterizing travel and service time uncertainties separately. The CSRI-based service-level constraints are integrated into both compact and set-partitioning models to mitigate the cascading effect. We further devise an exact branch-price-and-cut framework and employ a variable neighborhood search metaheuristic to achieve fast-effective solutions. Our numerical experiments with benchmark and real-world datasets validate the efficacy of our algorithms, underscore the superiority of adopting our systematic cascading delay mitigation strategy, and provide insights to AHH service providers about the topological structure and crucial tuning parameters. The CSRI and the dedicated solution methods can effectively support decision-making and enhance the punctuality of AHH services, leading to better service dependability and heightened stakeholder satisfaction.

3 - Emergency Department Overcrowding: Introducing an Innovative Virtual Triage System for Effective Demand Management

Mattia Cattaneo, University of Bergamo, Dalmine, Italy, Sebastian Birolini, Elisa Alessio, Filippo Manelli

The emergency department (ED) serves as a cornerstone of our society, offering around-the-clock emergency care. However, its effectiveness is hindered by issues like overcrowding, mainly fueled by a surge in "non-emergency" visits. Despite numerous attempts to tackle ED overcrowding, no single solution has fully resolved the issue. Yet, a promising solution is emerging: virtual triage. This technology allows remote assessment of patients, providing immediate triage recommendations to prevent overestimations, reduce unnecessary hospital admissions, and schedule healthcare demand.

Our focus lies in exploring how this cutting-edge remote solution can identify patients who, while not in urgent need, require specific ED investigations. By implementing an appointment-based service, we aim to enhance the overall patient experience, minimizing on-site wait times and optimizing resource utilization. To gauge the impact of this approach, we develop a simulation model mirroring current ED processes, assessing the new service's effect on ED performance. To this aim, we investigate patient acceptance through a stated preference analysis, where individuals express their preferences among three ED access alternatives: Traditional ED access, Scheduled ED access, Fee-scheduled ED access.

4 - Multi-Depot General Colored Traveling Salesman Problem in Home Healthcare System

ILGIN ACAR, Western Michigan University, Kalamazoo, MI, United States, Islam Altin

The home healthcare system is crucial especially for elderly, disabled people or those who are facing challenges in accessing healthcare services. Furthermore, extraordinary situations, such as pandemics, increase the importance of home healthcare services. Access to traditional healthcare is restricted as lockdowns make it impossible for people to leave their homes. Therefore, home healthcare system plays a critical role in facilitating access to healthcare services during these challenging times. Considering all these situations, this study focuses on the medication delivery problem in the home healthcare system. In this problem, multiple pharmacies situated in various locations can meet the medication demands of patients. In addition, while some patients' medication demands must be satisfied by specific pharmacies, others' medication demands can be met by any pharmacy. In this case, considering the service constraints, medication delivery must be completed within the specified time limit using the vehicles available at the pharmacies. Therefore, it is necessary to obtain optimal routes to minimize the total distance traveled by vehicles from different pharmacies. The problem encountered in medication delivery is addressed as the Multi-Depot General Colored Traveling Salesman Problem (MD-GCTSP). A mixed-integer mathematical model was developed to obtain an optimal solution for the MD-GCTSP. Furthermore, a metaheuristic algorithm was used to get quality solutions with short computing times for large-sized problems. The performance of the proposed methods is demonstrated on different-sized test problems. Computational results show that the proposed solution approaches are efficient and well-suited for solving the MD-GCTSP.

WB41

Summit - 433

Interacting Particle Systems and Queueing Systems

Invited Session

Applied Probability Society

Chair: Ruoyu Wu, Iowa State University, Ames, IA, United States

1 - Optimal Rate-Matrix Pruning for Large-Scale Heterogeneous Systems

Zhao Zhisheng, Georgia Institute of Technology, ATLANTA, GA, United States, Debankur Mukherjee

We present an analysis of large-scale load balancing systems, where the processing time distribution of tasks depends on both the task and server types. Our study focuses on the asymptotic regime, where the number of servers and task types tend to infinity in proportion. In heterogeneous environments, commonly used load balancing policies such as JFSQ/JFIQ exhibit poor performance and even shrink the stability region. Interestingly, prior to this work, finding a scalable policy with a provable performance guarantee in this setup remained an open question. To address this gap, we propose and analyze two asymptotically delay-optimal dynamic load balancing policies. The first policy efficiently reserves the processing capacity of each server for ``good'' tasks and routes tasks using the vanilla Join Idle Queue policy. The second policy, called the speed-priority policy, significantly increases the likelihood of assigning tasks to the respective ``good'' servers capable of processing them at high speeds. By leveraging a framework inspired by the graphon literature and employing the mean-field method and stochastic coupling arguments, we demonstrate that both policies achieve asymptotic zero queuing. Specifically, as the system scales, the probability of a typical task being assigned to an idle server approaches 1.

2 - Flocking: Stability, Chaos and Traveling Waves

Sayan Banerjee, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States, Amarjit Budhiraja, Dilshad Imon

We study a model for flocking given by a n-particle system under which each particle jumps forward by a random amount, independently sampled from a given distribution, with rate given by a non-increasing function w of its signed distance from the system center of mass. This model was introduced in Balázs et. al. (2014) and some of its properties were studied for the case when w is bounded. We are interested in the setting where w is unbounded which results in a stochastic dynamical system for interacting particles with fast and large jumps. We characterize the large n limit (the so-called `fluid limit') of the empirical measure process associated with the system and prove a propagation of chaos result. Next, for the centered n-particle system, by constructing suitable Lyapunov functions, we establish existence and uniqueness of stationary distributions and study their tail properties. In the special case where w is an exponential function and θ is an exponential distribution, by establishing that all stationary solutions of the McKean-Vlasov equation must be the unique fixed point of the equation, we prove a propagation of chaos result at t=∞ and establish convergence of the particle system, starting from stationarity, in the large n limit, to a traveling wave solution of the McKean-Vlasov equation. Our work answers several open problems posed in Balázs et. al.

3 - Limit Theorems for Heterogeneous Graphon Jsq(D) Model

Yan-Han Chen, Iowa State University, Ames, IA, United States, Arka Ghosh, Ruoyu Wu

We consider a variation of the supermarket model in which a task arriving at a dispatcher is routed to one of its neighborhood servers based on the Join-the-Shortest-Queue(d) policy. We consider both heterogeneous dispatchers and servers whose neighborhood relationships are described by a deterministic bipartite graph. The evolution of the queue length process for each server is described in the form of stochastic differential equations in which the interaction between servers exist. The law of large number results, both locally and globally, are established as the size of the system grows and the underlying graphs converge to a graphon. The limiting system is given by independent but heterogeneous processes governed by the graphon.

4 - Deep Backward and Galerkin Methods for the Finite State Master Equation

Cohen Asaf, University of Michigan, Ann Arbor, MI, United States, Mathieu Lauriere, Ethan Zell

This work proposes and analyzes two neural network methods to solve the master equation for finite-state mean field games (MFGs). Solving MFGs provides approximate Nash equilibria for stochastic, differential games with finite but large populations of agents. The master equation is a partial differential equation (PDE) whose solution characterizes MFG equilibria for any possible initial distribution. The first method we propose relies on backward induction in a time component while the second method directly tackles the PDE without discretizing time. For both approaches, we prove two types of results: there exist neural networks that make the algorithms' loss functions arbitrarily small and conversely, if the losses are small, then the neural networks are good approximations of the master equation's solution. We conclude the presentation with numerical experiments on benchmark problems from the literature up to dimension 15, and a comparison with solutions computed by a classical method for fixed initial distributions.

WB42

Summit - 434

Learning and Stochastic Control

Invited Session

Applied Probability Society

Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

Co-Chair: Hannah Li, Columbia, New York, NY, United States

1 - Inpatient Overflow Management with Proximal Policy Optimization

Jim Dai, Cornell University & CUHK-Shenzhen, Ithaca, NY, United States, Pengyi Shi, Jingjing SUN

Overflow patients to non-primary wards can effectively alleviate congestion in hospitals, while undesired overflow also leads to issues like mismatched service quality. Therefore, we need to trade-off between con gestion and undesired overflow. This overflow management problem is modeled as a discrete-time Markov Decision Process with large state and action space. To overcome the curse-of-dimensionality, we decompose the action at each time into a sequence of atomic actions and use an actor-critic algorithm, Proximal Policy Optimization (PPO), to update policy. Moreover, we tailor the design of neural network which represents policy to account for the daily periodic pattern of the system flows. Under hospital settings of different scales, the PPO policies consistently outperform some commonly used state-of-art policies significantly.

2 - Autodifferentiable Discrete Event Simulation for Queuing Network Control

Jing Dong, Columbia University, New York, NY, United States, Ethan Che, Hongseok Namkoong

Queuing network control is a ubiquitous task for managing congestion in job-processing systems, such as manufacturing processes, service systems, communication networks, healthcare delivery systems, etc. Recently, there has been a growing interest in applying reinforcement learning (RL) techniques to learn good control policies for general queueing networks. However, queuing network control poses some unique challenges, including (1) high stochasticity and non-stationarity, (2) large state and action spaces, and (3) lack of stability guarantees. In this work, we propose a scalable framework for policy optimization in general queuing networks based on auto-differentiable discrete-event simulation, which tackles the challenges listed above and substantially improves the sample efficiency over existing approaches. This framework is based on a novel approach for computing pathwise gradients for discrete event dynamical systems, where we address the non-differentiability of the dynamics through carefully designed smoothing. Our proposed method is easy to train (e.g., not sensitivity to hyper-parameter tuning), highly scalable (e.g., can solve large-scale problems), can handle more realistic instances outside the scope of previous RL methods, including systems operating in a non-stationary environment with non-exponential inputs (i.e., interarrival and service times).

3 - Stein's method for Markov processes: beyond exponentially distributed jump times

Anton Braverman, Northwestern University, Evanston, IL, United States, Ziv Scully

The generator comparison approach of Stein's method is well-known for being able to compare stationary distributions of Markov chains. To date, all applications of the generator comparison approach have been to continuous-time Markov chains where the jump times are driven by exponentially distributed clocks.

Using the virtual workload of the G/G/1 system as an example, I will showcase how the generator comparison approach can be applied to continuous-time Markov *processes* where inter-jump times need not have an exponential distribution, thereby expanding the scope of models where the approach can be applied.

4 - Statistical Inference for Markov Chains with Known Structure

Peter Glynn, Stanford University, Stanford, CA, United States, Lin Fan

In this talk, we will discuss nonparametric estimation for Markov chains that satisfy stochastic recursions of a known form, but in which the driving noise has an unknown distribution. We assume that we observe samples of the driving noise. We will argue that such Markov chains/data sets arise naturally within the setting of operations research and management science. We will discuss general statistical theory for such Markov chains, including consistency of “plug-in” estimators, the influence function, central limit theorems, and related computational issues. This work is joint with Lin Fan.

WB43

Summit - 435

Scheduling and Decision-Making in Stochastic Systems

Invited Session

Applied Probability Society

Chair: Ziv Scully, Cornell University, Ithaca, NY, United States

1 - Strongly Tail-Optimal Scheduling in the Light-Tailed M/G/1

George Yu, Cornell University, Ithaca, NY, United States, Ziv Scully

We study the problem of scheduling jobs in a queueing system, specifically an M/G/1 with light-tailed job sizes, to asymptotically optimize the response time tail. This means scheduling to make P[T>t], the chance a job's response time exceeds t, decay as quickly as possible in the t→∞ limit. For some time, the best known policy was First-Come First-Served (FCFS), which has an asymptotically exponential tail: P[T>t]~Ce−γt. FCFS achieves the optimal *decay rate* γ, but its *tail constant* C is suboptimal. Only recently have policies that improve upon FCFS's tail constant been discovered. But it is unknown what the optimal tail constant is, let alone what policy might achieve it.

In this paper, we derive a closed-form expression for the optimal tail constant C, and we introduce *γ-Boost*, a new policy that achieves this optimal tail constant. Roughly speaking, γ-Boost operates similarly to FCFS, but it pretends that small jobs arrive earlier than their true arrival times. This significantly reduces the response time of small jobs without unduly delaying large jobs, improving upon FCFS's tail constant by up to 50% with only moderate job size variability, with even larger improvements for higher variability. While these results are for systems with full job size information, we also introduce and analyze a version of γ-Boost that works in settings with partial job size information, showing it too achieves significant gains over FCFS. Finally, we show via simulation that γ-Boost has excellent practical performance.

2 - A Gittins Policy for Optimizing Tail Latency

Amit Harlev, Cornell University, Ithaca, NY, United States, George Yu, Ziv Scully

We consider the problem of scheduling to minimize the (asymptotic) tail latency in an M/G/1 queue. When the job size distribution is heavy-tailed, numerous policies have long been known to be asymptotically optimal (e.g. processor sharing, least attained service, SRPT). In contrast, for light-tailed distributions, only in the last few years have policies been developed that outperform FCFS. However, these policies all require job size information.

In this paper we find the optimal scheduling policy for minimizing asymptotic tail latency in the light-tailed M/G/1 with unknown job sizes. Surprisingly, the optimal policy turns out to be a novel variant of the Gittins policy, but with an unusual feature: it uses a negative discount rate.

3 - Economies of operational hybridity in large-scale service systems

Yunfang Yang, Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore, Singapore, Shuangchi He

Large-scale service systems often need to serve multiple classes of customers subject to distinctive delay requirements. For example, the inpatient department of a general hospital is usually required to make beds available within hours for patients transferred from the emergency department, whereas it may provide beds in days or weeks for elective patients. In this study, we consider a many-server queue with two classes of customers that are managed under two distinctive regimes: the waiting times of the first-class customers should be of a lower order of magnitude than those of the second-class customers (e.g., hours versus days). Besides servers dedicated to each class, there are cross-trained servers that can serve both classes. We first propose a sequencing policy for cross-trained servers to meet service-level requirements for the two classes with a minimal staffing level. Then, we identify the minimum number of cross-trained servers that is needed by the sequencing policy to minimize the overall staffing cost. We prove that under the proposed sequencing and cross-training scheme, the service system can achieve "economies of operational hybridity"—that is, the system can be dimensioned as if it operates under the more efficient regime that the second-class customers require, while a small fraction of cross-trained servers are sufficient to help fulfill the first-class customers' delay requirements.

4 - Cost-per-sample Bayesian optimization via the Pandora's box Gittins index

Qian Xie, Cornell University, Ithaca, NY, United States, Raul Astudillo, Peter Frazier, Ziv Scully, Alexander Terenin

Bayesian optimization is a technique for efficiently optimizing unknown functions in a black-box manner. To handle practical settings where gathering data requires use of finite resources, it is desirable to explicitly incorporate function evaluation costs into Bayesian optimization policies. This includes possibly stopping if the gains expected from gathering more data are insufficient to justify the cost of doing so. We introduce a new policy for cost-per-sample Bayesian optimization, using an unexplored connection with the Pandora's box problem, a well-studied problem from economics. The Pandora's box problem admits a Bayesian-optimal solution based on an expression called the Gittins index, which can be interpreted as an acquisition function. We demonstrate empirically that the same acquisition function performs well for Bayesian optimization, particularly in medium-high dimensions. The acquisition function is computationally efficient, and extends naturally from a uniform cost per sample to heterogeneous costs. Our work constitutes a first step towards integrating ideas from Gittins index theory into Bayesian optimization.

5 - Local hedging approximately solves Pandora's box problems with nonobligatory inspection

Ziv Scully, Cornell University, Ithaca, NY, United States, Laura Doval

We consider Pandora's box problems with nonobligatory inspection and single-item or combinatorial selection. A decision maker is presented with several items, each of which contains an unknown price, and can pay an inspection cost to observe an item's price before selecting it. Under single-item selection, the decision maker must select one item; under combinatorial selection, the decision maker must select a set of items that satisfies certain constraints. In our nonobligatory inspection setting, the decision maker can select items without first inspecting them. It is well-known that search with nonobligatory inspection is harder than the well-studied obligatory inspection case, for which the optimal policy for single-item selection (Weitzman, 1979) and approximation algorithms for combinatorial selection (Singla, 2018) are known.

We introduce a technique, *local hedging*, for constructing policies with good approximation ratios in the nonobligatory inspection setting. Specifically, local hedging takes as an input policies for the obligatory inspection setting and outputs policies for the nonobligatory inspection setting, at the cost of an extra factor in the approximation ratio. The factor is instance-dependent but is at most 4/3. We thus obtain the first approximation algorithms for a variety of combinatorial selection problems, including matroid basis, facility location, and prize-collecting Steiner tree.

WB44

Summit - 436

Advances in Medical Decision Support

Contributed Session

Chair: Anna K. Stirner, University of Cologne / Department of Business Administration and Health Care Management, Cologne, Germany

1 - An experiment on decision support in medicine

Anna K. Stirner, University of Cologne / Department of Business Administration and Health Care Management, Cologne, Germany, Bernhard Roth, Daniel Wiesen

Our experiment investigates how capacity constraints affect physicians' use of decision support systems to improve therapy decisions. Conducted with 247 German pediatricians, we manipulate physicians' capacity levels and the availability of decision support to observe changes in decisions regarding the length of antibiotic therapy for pediatric cases. Initially, physicians make decisions on therapy lengths for each case. Depending on their assigned experimental condition, subjects then either (i) decide whether to use decision support before making the final therapy decision, (ii) automatically receive decision support, or (iii) have no access to decision support. We vary information gathering costs, representing the capacity fraction needed for decision support. Subjects who decide or receive support can modify their initial therapy decisions for each case.

Results evidence that as capacity constraints increase, physicians' willingness to seek additional information declines. However, constraints did not affect the extent to which the gathered information was utilized. Importantly, our findings demonstrate that capacity constraints have a statistically significant and clinically relevant impact on the appropriateness of therapeutic decisions, affecting the quality of care, especially among less experienced physicians. The likelihood of gathering additional information increases significantly with case severity. Utilizing additional information leads to improvements in the appropriateness of therapy decisions in cases of undertreatment, while overtreatment persists even with decision support.

In conclusion, reducing capacity constraints could enhance the use of decision support, thus improving the appropriateness of therapeutic decisions. These findings have significant implications for healthcare management, suggesting that decreasing capacity constraints could benefit healthcare quality.

2 - Multi-Objective Optimization of Multi-Agent Kidney Transplantation Considering Fairness

Shayan Sharifi, Wayne State University, Detroit, MI, United States, Evrim Dalkiran

Despite advancements in living donor transplantation and kidney exchange programs, the demand for kidney transplants far exceeds the available supply, leaving thousands on waiting lists. This study presents a comprehensive Multi-Agent Kidney Exchange Programs (MKEP) model aimed at optimizing transplant numbers and enhancing transplant quality within the U.S. healthcare system.
In this paper, we introduce a novel MKEP model designed not only to increase the number of transplants available to patients but also to improve the quality of these transplants, thereby enhancing patients' post-transplant quality of life. Furthermore, we classify existing patients into three groups: normal, sensitive, and urgent, enabling us to prioritize kidney transplants for emergency patients. We utilize a multi-objective model to balance transplant quality and delay time, establishing an acceptable lower bound for human leukocyte antigens (HLA). Our findings demonstrate that employing this approach enhances both the quantity and quality of transplantations compared to traditional methods.

3 - Improving Outcome Predictions in Chronic Disease Management: A Case Study on Chronic Kidney Disease Using Multi-Sourcing, Integrated Data

yubo li, Carnegie Mellon University, Pittsburgh, PA, United States, Rema Padman

This study presents a novel predictive framework that integrates clinical and administrative claims data, leveraging advanced machine learning and deep learning techniques to forecast Chronic Kidney Disease (CKD) progression to End-Stage Renal Disease (ESRD). Our multi-stage approach encompasses data preprocessing, model development, and comprehensive analysis, employing a diverse set of models including logistic regression, random forest, XGBoost, and deep learning architectures such as temporal convolutional networks (TCN) and long short-term memory (LSTM) networks. The results demonstrate the superiority of deep learning models, particularly LSTM, in capturing the temporal dynamics of CKD progression, with a 24-month observation window emerging as the most effective timeframe for ESRD prediction across all data types. The application of explainable AI techniques provides insights into the complex interplay of predictive features, while the incorporation of a revised, race-neutral eGFR equation addresses potential bias and improves predictive accuracy for African American patients. Despite certain limitations, this study establishes a robust foundation for early detection and personalized risk assessment of ESRD, highlighting the power of integrating diverse data sources, advanced modeling techniques, and explainable AI in predicting CKD progression to ESRD.

4 - Patient Pathways and Team-Based Care for Cancer Patients in Kuwait: Process Mining Approach

Abdullah Alibrahim, Kuwait University, Kuwait, Kuwait, Munirah Alateeqi, Abdulwahab Al-Tourah, Waddah Al-Refaie

Cancer care demands systems that support multidisciplinary team-based care. This study aims to analyze patient pathways and team-based care metrics by examining various specialties engagement and how processes are standardized. We utilized linked two datasets: national-level discharge data from Kuwait Ministry of Health (MOH) hospitals and patient records from Kuwait Cancer Control Center (KCC). For this analysis, 50 cancer patients were selected from the two most prevalent types: breast and colon cancer, capturing 2,182 and 2,507 patient activities at KCC and hospitalizations from MOH discharge data for each cancer type, respectively. Our analysis uncovered significant variability in treatment patterns across different care stages, including care initiation, therapy, and surveillance. The analysis indicates that colon cancer patients experience more per-patient events than breast cancer patients during a longer treatment journey length (51.4 months for colon cancer vs. 49.6 months for breast cancer patients). Breast cancer patients required balanced therapeutic modalities with systemic treatment, radiation, and surgery. In contrast, colon cancer patients primarily underwent outpatient visits and medical infusions. Furthermore, 24% of colon cancer patients' hospitalization episodes occurred in general hospitals outside of KCC, indicating an urgent need for coordination and medical information sharing with other treatment facilities. Breast cancer patients often started treatment after registration, with 68% initiating treatment this way and the rest through a rapid-screening pathway. In contrast, all colon cancer patients were registered before treatment at KCC. These findings emphasize the importance of departmental coordination and standardization with a focus on conformance and enhanced personalized journeys.

WB45

Summit - 437

New Advances in Organ Transplantation Modeling

Invited Session

Health Applications Society

Chair: Amirhossein Moosavi

Co-Chair: Yili Wang, University of Michigan, Ann Arbor, MI, United States

1 - Optimizing Sequential Decisions in Marginal Kidney Acceptance

Yili Wang, University of Michigan, Ann Arbor, MI, United States, Amirhossein Moosavi, Mariel Lavieri, Danielle Haakinson, Yee Lu, Kate Giles

Chronic kidney disease affects over 37 million people in the United States, with more than 700,000 patients requiring renal replacement therapy, including dialysis or kidney transplantation. Despite nearly 100,000 patients on the US transplantation waitlist, about 20% of lower-quality (marginal) kidneys are discarded annually. This ongoing project aims to develop models that assist patients and clinicians in making informed decisions about accepting marginal kidneys while considering their risk aversion level. We will present our pilot model and explore potential future research directions.

2 - A Data-Driven Approach for Identifying Candidates for Xenotransplant Human Clinical Trials

NAZ YETIMOGLU, The University of Chicago Booth School of Business, Chicago, IL, United States, Baris Ata, Robert A. Montgomery, Jesse D. Schold

The demand for transplant organs surpasses supply, with xenotransplantation offering a potential solution to this shortage. The successful investigational transplants of genetically edited pig kidneys into brain-dead patients and two, single patient IND enabled xeno-heart transplants into living human recipients signal that the first human clinical trials are imminent. Using the 2-year survival of non-human primates with gene-edited pig kidneys in pre-clinical studies as a benchmark, we developed a tool that identifies patients predicted to have a shorter life expectancy in the current system than with a xenotransplant. Using DeepSurv, Random Survival Forest and Cox Proportional-Hazards models, we found that it is hard to reach clinical equipoise unless the expected xenograft survival exceeds two years. As such, it is difficult to identify many candidates who would benefit from trials and potential beneficiaries are spread across 200 transplant centers. However, several incentives could allow more patients to reach equipoise. Keeping patients inactive on the waitlist while they have a functioning xeno-kidney incentivizes them, slightly improving equipoise. Giving patients with failed xeno-kidneys the same priority as prior living donors moves the needle towards equipoise marginally. Lastly, giving highest priority when a xeno-kidney fails could achieve equipoise for substantial number of patients.

3 - Thousands Unused: The Problem of Lossy Allocation in Kidney Transplantation and Opportunities for Operations Research

Elijah Pivo, MIT, Cambridge, MA, United States, Maura Hegarty, Theodore Papalexopolous, Nikolaos Trichakis, Dimitris Bertsimas

While tens of thousands hope to receive a kidney transplant, thousands of donated kidneys go unused annually. The vast majority are not accepted by a transplant candidate during the allocation process. In recent years, the nonuse rate has grown to greater than 25% of donated kidneys. We discuss potential clinical and logistical drivers for this phenomenon, as well as opportunities for machine learning and optimization to improve the allocation process.

4 - Personalized Data-Driven Model for Organ Acceptance

Jiahui "Gary" Luo, Dartmouth College, Hanover, NH, United States, Wesley Marrero, Mariel Lavieri, David Hutton, Neehar Parikh, Lawrence An

Our method for creating a personalized data-driven model for organ acceptance incorporates machine learning, survival analysis, and mathematical optimization. In this work, we focus on liver transplants, which are the second most common transplant in the United States. We employ machine learning-based methods to predict liver quality and forecast patient survival rates. Based on these predictions, we develop a Markov decision process (MDP) framework to represent the decision-making process of accepting or declining organ offers tailored to an individual candidate on the liver transplantation waiting list.

WB46

Summit - 438

Decision-Making in Healthcare Systems

Invited Session

Health Applications Society

Chair: Zeyu Liu, West Virginia University, Morgantown, WV, United States

1 - Personalized and Interpretable Screening Policy for Diabetic Eye Disease

Taewoo Lee, Seoul National University, Seoul, Korea, Republic of, Poria Dorali, Soroush Akbarijokar, Christina Weng

Diabetic retinopathy (DR) is the leading cause of blindness among working-age Americans. Currently only 30-60% of patients adhere to annual screening guidelines due to limited access to eye care, particularly impacting medically underserved populations. Teleretinal imaging

(TRI) is emerging as an affordable tool that has the potential to increase screening rates for those with limited access. We develop a partially-observable Markov decision process (POMDP) that uses various patient-specific characteristics and determines personalized screening policies for TRI-based screening. To promote clinical implementability we then convert the POMDP-driven policy into a more interpretable screening policy. Finally we analyze the cost-effectiveness of the proposed policies against current screening guidelines.

2 - Dynamic Medical Countermeasure Deployment Under Epidemic Outbreaks

Mohsen Mehrabiyan, West Virginia University, Morgantown, WV, United States, Zeyu Liu

In recent years, medical countermeasure (MCM) deployment has been increasingly investigated to mitigate the impact of infectious disease outbreaks and enhance the preparedness and readiness of the national healthcare system. To address the lack of robust health security and the insufficiency in emergency response capabilities, there's still a critical need for an effective method to optimize MCM deployment strategies facing the complex dynamics during epidemics. In this study, we utilize a multi-stage stochastic programming (MSP) model to optimize emergency supply planning, including resource allocation and facility location decisions that optimize the distribution of MCMs, as well as sequential decision-making in response to disruptive uncertainties. In our model, we prioritize actions that ensure fair access to healthcare resources during emergencies, with a particular focus on supporting marginalized groups. Because of the computational difficulties arising from the MSP tree structure, we design novel computational algorithms to produce high-quality solutions efficiently. We collect real-world data from literature and design agent-based simulations for scenario realizations that quantify uncertainties according to individual characteristics. Through carefully designed experiments, our model enables comprehensive analysis of disease transmission, healthcare delivery, and public sentiment during infectious disease outbreaks.

3 - Equitable Anesthesiologist Scheduling Under Demand Uncertainty Using Multi-Objective Programming

Kai Sun, University of Texas at San Antonio, San Antonio, TX, United States, Minghe Sun, Ronald Dravenstott, Frank Rosinia, Arkajyoti Roy

This work addresses a practical anesthesiologist scheduling (AS) problem under clinical demand uncertainty. The goal is to plan and deploy providers to meet clinical demand and institutional protocols. A data-driven two-step sequential AS framework is developed using mixed-integer multi-objective programs. Step 1 designs shifts to minimize excessive clinical time using conditional value-at-risk constraints to account for demand uncertainty. Step 2 assigns shifts to providers considering optimal and equitable workload distribution and the number of required providers. An ε-constraint solution method is applied for multi-objective optimization, and an iterative solution method is developed to improve workload equity in clinical applications. Two case studies, the budget and hiring planning and the monthly AS, are addressed in a large anesthesiology department via the AS framework. For budget and hiring, decision makers can make tradeoffs based on their preference using the nondominated frontiers obtained via the ε-constraint method. For monthly AS, the iterative solution method can capture institutional requirements, accommodate pre-assigned shifts when incorporating complex clinical and lifestyle preferences, and further improve workload equity. The workload variance has been substantially reduced by 52.3% after the implementation based on the historical schedule data. The provider schedule satisfaction has been improved from 3.13/5 to 3.47/5, and at least 82% of scheduling burdens on department leaders have been relieved. The developed framework has been deployed and has become the key component of operating room scheduling in the partner hospitals, including University Hospital and VA Hospital, in San Antonio.

4 - Interpretable Deep Learning for Diabetic Retinopathy Prediction Using Electronic Health Records

Jiahao Shao, University of Tennessee, Knoxville, Knoxville, TN, United States, Zekai Wang, Anahita Khojandi, Tieming Liu, Bing Yao

Diabetic Retinopathy (DR), a complication of diabetes affecting the small blood vessels, stands as the primary reason for vision impairment in adults of working age. This condition often advances to critical stages due to inadequate screening participation and the prohibitive costs of necessary diagnostic technology. However, the widespread availability of Electronic Health Records (EHR) offers a significant opportunity to overcome these challenges through the use of machine learning techniques. This study introduces an interpretable, non-image-based deep learning model designed for data-driven detection of DR. By leveraging the Attentive Interpretable Tabular Learning (TabNet) architecture, along with employing missing value masks and multi-branching technique, this model aims to tackle the prevalent issues of missing value and data imbalance issues in medical datasets for accurately DR detection. Additionally, unlike traditional deep neural networks, which suffer from a lack of interpretability, our model is capable of providing the importance of each medical variable, enhancing transparency and usability for healthcare providers. This innovative approach has the potential to enable confident recommendations for personalized ophthalmic exams and screening frequencies, particularly targeting at-risk patients.

WB47

Summit - 439

Optimizing Patient Flow and Readmission Risk in Healthcare Systems

Contributed Session

Chair: Suman Mallik, University of Kansas

1 - An Interpretable Imbalance Ensemble Classification Method for Readmission Risk Assessment Incorporating Multi-view Perturbation and SHAP Analysis

Shaoze Cui, Beijing Institute of Technology, Beijing, China, People's Republic of, Ruize Gao, junwei kuang, Liang Yang, Huaxin Qiu

In the realm of medical services, patients may experience readmissions shortly after discharge due to unreasonable discharge arrangements or recurring illnesses. This not only wastes valuable medical resources but also diminishes patient satisfaction. To this end, we propose an interpretable imbalance ensemble classification method incorporating multi-view perturbation to assess the readmission risk of inpatients. Compared to existing research, this study designs a novel multi-view perturbation method to enhance the model's generalization performance. Moreover, a more fault-tolerant ensemble strategy based on evidential reasoning (ER) rules is proposed, which makes the fusion results of the ensemble learning model more stable. Meanwhile, considering that sensitive parameters can affect model performance, we propose a parameter optimization method based on the differential evolution algorithm, which integrates model predictive performance and computational efficiency when designing the fitness function. Experimental results using real-world medical data demonstrate that the proposed method effectively identifies patients with high readmission risk and outperforms existing state-of-the-art readmission risk assessment methods.

2 - Optimizing Early Discharge: Trade-Offs between Capacity and Readmissions

Zhiyuan Lou, Technical University of Munich, Heilbronn, Germany, Taozeng Zhu, Jingui Xie, Gar Goei Loke

In this work, we consider the ward capacity management problem with readmissions, where the decision-maker optimizes the elective schedule and early discharge policy, so as to minimize bed shortages. Existing research has shown that early discharge can lead to higher rates of readmission, and longer readmission length-of-stay. This sets up the need to balance the temporal trade-off between the immediate capacity freed up by early discharges and increased readmissions down the road. Such re-entry structure creates challenges when modelling via traditional methods. We appeal to the Pipeline Queues (Bandi and Loke 2018) framework, and propose an optimization model where the early discharge policy is expressed as a state-dependent decision rule. The model has a reformulation, which can be solved as a sequence of convex programs with asymptotically linear constraints. In our numerical study, we identify an intermediate region of the probability of readmissions where time-invariant policies can lead to as much as 77% more shortages. Ignoring the effects of early discharge on readmissions can lead to at least 75% and 150% more bed shortages in time-homogeneous and non-time-homogeneous settings respectively, even against un-optimized elective admissions. Using optimal early discharge strategies without jointly optimizing elective admissions will lead to 20% more shortages.

3 - Developing a decision support system to predict delayed discharge from hospitals using machine learning

Majid Taghavi, Saint Mary's University, Halifax, NS, Canada, Mahsa Pahlevani, Peter Vanberkel, Enayat Rajabi

The increasing demand for healthcare services poses significant challenges in effectively managing patient flow, particularly concerning patients classified as Alternative Level of Care (ALC). These patients, although no longer in need of acute care, often encounter obstacles to discharge and cause several issues, such as hospital overcrowding and compromised health outcomes. Using administrative health data from Canadian hospitals, this study develops machine learning models to identify potential ALC patients and estimate their hospital length of stay as early as their admission time. Our findings show the efficacy of the eXtreme Gradient Boosting algorithm in accurately predicting potential ALC patients, while the Random Forest regression model surpasses others in forecasting the length of stay for ALC patients. To understand how the predictions are made, an explainable machine learning approach (SHAP) was used to identify the most important features that impact machine learning outcomes. We developed two sets of user-friendly and easy-to-follow guidelines for hospital staff to identify ALC patients proactively, estimate their length of stay, and mitigate the patient flow challenges posed by the ALC patients. Hospitals can use such decision-making tools to optimize resource allocation, enhance operational efficiency, and ultimately improve patient care outcomes.

4 - The Interactions of Crowding, Patient Severity, and Queue Rank at a Hospital Emergency Department

Suman Mallik, University of Kansas, Lawrence, KS, United States, Lu Wang, Mazhar Arikan

This study empirically investigates the impacts of additional patient arrivals and changing queue rankings on patient processing in the emergency department (ED). Through causal mediation analysis, we demonstrate that changes in queue rank mediate the effect of additional arrivals of higher-severity patients on patient waiting time. However, additional arrivals of patients with the same or lower severity levels directly affect patient waiting time.

WB48

Summit - 440

Learning and Experimentation in Markets and Platforms

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Shukai Li, Northwestern, Evanston, IL, qwer1234, United States

1 - Semi-Nonparametric Adaptive Bidding in Repeated First-Price Auctions Under Binary Feedback

Yongyi Guo, University of Wisconsin Madison, Madison, WI, United States, Danqi Luo, Hongyu Shan, Zhengyuan Zhou

In the past three years, a particularly important development in the digital advertising
industry is the shift from second-price auctions to first-price auctions for online display ads. This shift immediately motivated the intellectually challenging question of how to bid in first-price auctions, because unlike in second-price auctions, bidding one's private value truthfully is no longer optimal. In this paper, we study how to adaptively bid in repeated first-price auctions under binary feedback–currently used by almost all ad exchanges–where the bidder only sees whether he wins or loses the bid after each auction. We develop a semi-nonparametric adaptive bidding policy that achieves low regret. Our results reveal an interesting phenomenon: as the non-parametric uncertainty component becomes smoother,
the regret performance also improve. In the limit, when the component becomes infinitely smooth, we obtain minimax regret up to log factors.

2 - Learning User Behavior in a Social Network with Limited Data

yunduan lin, University of California, Berkeley, Berkeley, CA, United States, Haoting ZHANG, Zeyu Zheng, Zuo-Jun Shen

Learning user behavior in a social network involves the understanding of both individual characteristics and the network effects arising from interactions among individuals. There are primarily two challenges. Firstly, the network diffusion process is dynamic and governed by intricate network structures. This complexity often makes the outcomes of diffusion lack straightforward, closed-form expressions. Secondly, calibrating network diffusion models is complicated by limitations in available data, both temporally and spatially. From the temporal aspect, capturing its entire trajectory of network diffusion may require extensive efforts and thus is often not feasible. From the spatial aspect, network structures are typically specific and limited. While individual-level data might be abundant, sampling network-level structures as desired is usually not possible. In this work, we aim to present a comprehensive framework for learning user behavior in a social network based on one single observed instance. We first analyze the learning performance under different data availability using a specific diffusion model. Then, we employ simulation as a vital tool to establish the framework for a broad class of models. Specifically, simulation enables the modeling of complex systems and the observation of their behavior under various conditions. By constructing a simulation model of the network, we can generate output data corresponding to different parameter values. We compare this simulated data to the observed data in an attempt to identify the values that yield the closest match by the definition of an objective function, which quantifies the distance between the given data and the simulated data.

3 - Value of One Data Point: Active Label Acquisition in Personalized Assortment Optimization

Mo Liu, University of North Carolina - Chapel Hill, Chapel Hill, NC, United States, Junyu Cao, Zuo-Jun Shen

Predicting customers' preferences based on their features is crucial for personalized assortment optimization. When building this prediction model, using informative data can significantly increase the expected revenue from personalized assortments. This paper studies how to sequentially collect informative data to construct this prediction model. We introduce a novel concept, the 'value of one data point,' which evaluates the marginal contribution of acquiring a specific customer's preference to the expected revenue in personalized assortment optimization, given the existing training set. Notably, this value drops to zero once the optimal assortment for this specific customer is determined. To estimate this value and identify important customers for acquiring their preferences, we derive a feature-dependent upper bound. This bound provides significant insights into the importance of each data point for revenue growth. Based on this upper bound, we develop a personalized incentive policy for effectively collecting survey data from customers to obtain their preferences. We provide non-asymptotic guarantees for both the cumulative incentives and the revenue from the final prediction model. Theoretically, we show that our personalized incentive policy requires smaller cumulative incentives than any fixed incentive policy to achieve the same level of revenue. Additionally, our numerical experiments with real-world and synthetic datasets validate the effectiveness of our personalized incentive algorithms over fixed strategies.

4 - Sequential Decision-Making in Online Experimentation: When Safety Meets Efficiency

Feng Zhu, Massachusetts Institute of Technology, Cambridge, MA, United States, David Simchi-Levi

This talk explores the integration of safety into efficiency in sequential decision-making for online experimentation, specifically within a stochastic multi-armed bandit setting. While previous work mostly focuses on achieving efficiency by minimizing regret expectation, controlling the regret tail risk to ensure safety is essential in applications such as revenue management, clinical trials, and financial investment, but has not been well studied. This work provides a detailed characterization of regret distribution under the safety concern of managing tail risk. In Part I, we aim to design policies that enjoy both optimal regret expectation and light-tailed regret distribution. We first find that any policy that obtains the optimal instance-dependent regret expectation could incur a heavy-tailed regret distribution. We then design a novel policy that enjoys the optimal worst-case regret expectation and has the optimal worst-case regret tail risk with an exponential decaying rate. Our policy design also bears an interesting connection with Monte Carlo Tree Search (MCTS) used in AlphaGo. In Part II, we study the optimal trade-off between expectation and tail risk for regret distribution. Our results reveal several managerial insights on how to design policies that balance efficiency and safety. All our results can be extended to the stochastic linear bandit setting.

5 - Atlanta Police Zone Redesign for Neighborhood Integrity

Zheng Dong, Georgia Institute of Technology, Atlanta, GA, United States, Yao Xie

Atlanta, the capital of the U.S. state of Georgia, plays an important role as the cultural and economic center of the Atlanta metropolitan area, which is home to 6.3 million people and the sixth-largest metropolitan area in the United States. The police departments in Atlanta (APD) have long struggled with staffing shortages. Officers, facing longer shifts and broader coverage areas, must strategically allocate their efforts. Georgia Tech (GT) has worked with Atlanta Police in 2019 to redesign zones and balance workload. However, socioeconomic development since COVID-19 pandemic changes the crime landscape increase again the imbalance of the police workload across Atlanta city area, calling for new police zone design. In this work, we collect data during post-pandemic era and re-design the police zones in Atlanta considering neighborhood integrity. We also use new metrics to balance local workload of officers between zones when dealing with 911 calls.

WB49

Summit - 441

Empirical Research on Sustainability in the Auto Industry and Supply Chains

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Iman Attari, Indiana University Bloomington, Bloomington, IN, United States

Co-Chair: Jonathan Helm, Indiana University, Bloomington, IN, 47405, United States

1 - Impact of Franchise Car Laws on EV Adoption

Andrey Makhanov, UNC Kenan-Flagler Business School, Chapel Hill, NC, United States, Nur Sunar, Jay Swaminathan

We empirically study the effect of the auto franchise law on electric vehicle (EV) adoption.

2 - Regulatory Power in Automotive Product Recall Effectiveness: An Empirical Investigation

Ramin Sepehrirad, Washington State University, PULLMAN, WA, United States, Kevin Mayo, George Ball

While significant research has explored causes and consequences of product recalls, little is known about recall effectiveness: how well a recalling firm performs in repairing faulty products. Anecdotal reports indicate that firms are relatively ineffective in this endeavor, leading to millions of recalled products remaining in the market, unaddressed. We study how regulatory authority in automotive product recalls impacts firms' remedial performance. Drawing upon Ocasio's attention-based view (ABV) and extant recall literature, we develop competing hypotheses of how mandating a recall affects repair rates — the proportion of remedied to affected vehicles. Using a unique dataset containing more than 4200 recalls on passenger vehicles from 2009 to 2022, we surprisingly find that when the regulator mandates a recall, fewer vehicles are repaired eventually. Specifically, in a quasi-experimental setting, we show that mandating reduces repair rates at least by 5%, equivalent to 20,000 vehicles for a typical recall campaign. We further investigate two moderation mechanisms and find that mandating negative effects is stronger when the estimated repair time is longer, or when the recall campaign is more costly. Finally, we examine how potential hazards of the defective component in a recall campaign impact remedial procedure effectiveness. Theoretical contributions and practical implications are discussed.

3 - Shifting Attention and the Green Ripple Effect: Unveiling Green Innovation Diffusions in Supply Chain Networks

Nima Safaei, University of Illinois Urbana-Champaign, Champaign, IL, United States, Jafar Namdar, Gautam Pant, Shagun Pant

In addition to internal R&D motivators, peers' innovation is an important stimulus to the focal firm's innovation performance. While it is well-known that firms in vertical relationships within supply chain networks affect each other's strategic decisions, operations, and innovation, our study focuses on an unexplored area: the upstream diffusion of attention shift toward more green innovation in the global supply chain network. We investigate whether and how (i.e., the mechanism) customers' attention shifts toward more green innovations, captured through patents, diffuse to suppliers. We further study the relationship between such an attention shift and firms' environmental performance and evaluate the impacts on firms' financial performance. We leverage a comprehensive dataset of 418,219 global supply chain linkages, combined with patent information and performance metrics of the firms in the relationship, and employ a structural econometric model to identify the diffusion effect. We find that shifting to more green innovation by customers directly diffuses to suppliers; the peer influence channel (i.e., the direct impact of customers' green innovation ratio on the green innovation ratio of the focal firm) is by far the most significant source of diffusion. We also find a positive impact of increased green innovation ratio on firms' environmental and financial performance, highlighting the critical role of focusing on more green innovation in achieving sustainable business practices and economic benefits.

WB50

Summit - 442

Emerging Topics in Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Fernanda Bravo, UCLA Anderson School of Management, Los Angeles, CA, United States

Co-Chair: Maria Ibanez, Kellogg School of Management at Northwestern University, Evanston, IL, United States

1 - Mo-dalities, mo problems: The hidden costs of hybrid work in primary care practices

Mitchell Tang, Harvard Business School, Boston, MA, United States, A Jay Holmgren, Robert Huckman, J Michael McWilliams, Maximilian Pany

Many workers today have hybrid work schedules with a mix of in-person and virtual meetings. While the ability to conduct meetings virtually undoubtably has benefits, when integrated into a still predominately in-person schedule, it also introduces new frictions. Transitions between in-person and virtual meetings within the same day can burden workers. On the other hand, while creating dedicated virtual work periods reduces the likelihood of these transitions, it can also reduce capacity utilization if there isn't sufficient demand for virtual meetings in that specific window.

We examine these new frictions in the context of hybrid primary care practices, which offer both in-person and telemedicine (virtual) visits. We observe large negative impacts specifically for transitions from in-person to telemedicine visits. Subsequent telemedicine visits saw a 19% lower likelihood of on time start and patients were 33% more likely to abandon the visit before being seen. These disruptions also manifested in fewer logged diagnoses, lower rates of timely note completion, and lower patient-reported provider ratings. However, when providers instead allocated dedicated appointment slots for telemedicine, those slots saw an 11% reduction in booking rate. The negative impacts of dedicated telemedicine appointment slots were lessened for time slots and weeks when patient telemedicine demand was higher.

Our findings demonstrate the subtle costs of hybrid work. They also demonstrate the tradeoffs of incorporating explicit structure in a hybrid schedule, the types of work that may be better suited for dedicated virtual blocks, and the value of synchronizing schedules to minimize loss from those blocks.

2 - Reducing Appointment Delays: The Impact of Standardized EHR Usage on Physician Timeliness

Bradley Staats, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Umit Celik, Sandeep Rath, Saravanan Kesavan

Problem Definition: This study explores the impact of standardized documentation tasks within Electronic Health Record (EHR) systems on physician timeliness. By analyzing data from physicians' schedules and detailed EHR timestamps, we assess how standardizing EHR tasks before and during appointments affects appointment lengths and delays.

Methodology/Results: Utilizing classical propensity score matching and machine learning for causal inference, our analysis reveals that standardizing EHR tasks either before or during appointments significantly reduces the likelihood of delays in subsequent appointments. The primary mechanism identified is the shortening of appointment durations through EHR standardization. While increased standardization generally leads to shorter appointments, it does come with trade-offs, notably longer subsequent appointments for the same patient due to unnecessarily prolonged notes generated by standardization tools.

Managerial Implications: The findings underscore the importance of EHR standardization in enhancing physician timeliness and patient satisfaction. By mitigating appointment delays, physicians can potentially increase patient throughput and improve the quality of medical services provided. This study provides key insights for healthcare administrators seeking to optimize operational efficiency and patient care through strategic EHR standardization.

3 - Incentivizing Drug Rediscovery: Patent Extension vs. Accelerated Regulatory Pathways for Repurposed Therapeutics

Fernanda Bravo, UCLA Anderson School of Management, Los Angeles, CA, United States, Elodie Adida

Drug repurposing, involving the use of an already approved drug for a newly identified treatment, is seen as a method to expedite access to new therapies at a reduced cost. Despite its advantages compared to de novo drug development, drug repurposing is believed to be underutilized. We examine strategies for encouraging drug repurposing. While providing financial incentives to drug manufacturers can facilitate this process, we must carefully weigh the potential impact on patient well-being. We explore two approaches: extending exclusivity and expediting reviews, preventing cross-labeling between indications, and prohibiting narrow labeling. We assess the optimal design of each approach and its effects on patient welfare.

4 - When life gives you lemons: Spillovers between pharmaceutical recalls and shortages

Hanu Tyagi, University of Illinois, Urbana-Champaign, IL, United States, Rachna Shah, Junghee Lee

Operational failures in the pharmaceutical industry occur frequently and prolong for a long time. Hence, there may be spillovers between failures, i.e., the onset or recovery of one type of failure could impact the occurrence and recovery time of other types of failures. We investigate such spillovers and ask: How does a firm's recall experience impact shortages? Using recalls and shortages data for 319 firms from 2010 to 2019, we show that completed recall experience reduces the likelihood of future shortages suggesting internalized learning, whereas ongoing recall experience prolongs the shortage duration and shortage likelihood indicating operational burden.

WB51

Summit - 443

Empirical Research in Operations Management

Invited Session

MSOM: Service Operations

Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Yuan Ma, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

1 - Health on Loan: The Effect of Local Credit Availability on Hospital (Re)Admissions

Yuan Ma, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Bank loans are crucial sources of finance for hospitals. This paper assesses the effect of increases in local credit supply on healthcare outcomes. We show that the total admission drop when there are more credits available in local banks, and the decrease is driven by decreases in readmission.

2 - The Impact of Senior Living Facilities on Healthcare Utilization

Carri Chan, Columbia Business School, New York, NY, United States, Ann Bartel, Minje Park, Fanyin Zheng

The senior population in the U.S. has been rapidly growing. Consequently, enrollment and spending in Medicare, the federal health insurance program for individuals aged 65 and older, have skyrocketed. At the same time, senior living facilities (SLFs) -- assisted living facilities and independent living facilities -- have drastically expanded in the U.S. to serve the needs of the elderly. SLFs provide housing in a community environment that often includes personal care assistance services. Thus far, there has been no research studying the effects of SLFs on healthcare utilization. We utilize data of Medicare claims and SLF real estate development to shed light on this question. Although we have granular data, analyzing the causal relationship with a naive regression will likely yield a biased estimate due to endogeneities in the decision to build SLFs in certain areas. To address potential endogeneities, we use the instrumental variables approach with a novel instrument variable motivated by the real estate and economics literature. Our analysis finds that senior living facilities significantly reduce Medicare utilization (both the dollar spent and length of stay) for skilled nursing facilities, the short-term, post-acute care provider where patients go after getting discharged from a hospital to continue recovering after an illness, injury, or surgery.

WB52

Summit - 444

Optimization for Natural Resource Extraction and Processing

Invited Session

Energy: Natural Resources

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, 80401, United States

1 - Improving the Effectiveness of Large-Scale Data Analysis Through Multiple Regression

Edikan Udofia, Colorado School of MInes, Golden, CO, United States

This project explores predictive modeling of roll forces in hot rolling, crucial for steel manufacturing efficiency and product quality. Using a dataset from a steel mill, it applies multiple regression, variable transformations, and selection techniques to improve prediction accuracy. Incorporating strain rate effects and addressing variable conversion and collinearity, the study enhances both predictive modeling and the understanding of hot rolling dynamics.

2 - Optimizing Haul Truck Maintenance Scheduling in Open Pit Mines

Bart Maciszewski, Imperial Oil, Calgary, AB, Canada, Jose S Rodriguez, Anuraj Grewal, Christine Viljoen, Amber McKay, Ervin Balleza, Andrew Milne, Jacob Olafsen, Cathy Giang, Hussam Mansour, Pujit Jetley, Brendan Cumming

We present a problem of scheduling maintenance for heavy haul trucks in an oil sands mine. The trucks require periodic preventive maintenance based on operating hours and in addition to break-in work. There is a need to schedule close to due dates while grouping activities to minimize the number of truck visits to the shop. Activities need to be performed on compatible bays and need to respect the amount of available resources. We decompose the problem into a grouping and detailed scheduling stage. We model grouping as a general assignment problem and scheduling as a special case of the Dual Constrained Hybrid Flow Shop with Skipped Stages. We solve both stages using constraint programming with rolling horizon decomposition for the scheduling stage. We show how the model has uncovered important business insights and improves over current approaches.

3 - Optimization of Daily Casting Schedule in Steel Production Using Mixed-Integer Programming

Jenifer McClary, Colorado School of Mines, Golden, CO, United States, Hayden Archer, John Cox, Alexandra Newman, Brian Thomas

This study presents an advanced mixed-integer programming model tailored for optimizing the daily casting schedule within a steel mill. The model addresses the intricate processes involved in alloying, refining, casting, and rolling of molten steel with the primary objective of minimizing penalties incurred due to deviations from mill best practices while adhering strictly to safety and logical constraints. Leveraging cutting-edge software, we solve large instances containing hundreds of variables and thousands of constraints. Results indicate potential cost reductions of up to 40%, thereby enhancing operational efficiency and reducing manufacturing costs in the steel industry.

4 - Trade-Offs Between Equity and Revenue in Electric Vehicle Charging Station Siting and Sizing

Felix Ayaburi, Colorado School of Mines, Golden, CO, United States, Yongchen Wu

With the increase in electric vehicle purchase for private passenger use, we focus on optimizing electric vehicle charging infrastructure across the United States. Specifically, while developers possess the incentive to place public charging stations in urban locations that will garner them the most revenue, underserved communities and neighborhoods, both commercial and residential, will be left without the opportunities to charge their vehicles, thus hindering adoption and resulting in increased exposure to emissions and pollution relative to more affluent neighborhoods. In turn, this impedes progress towards sustainable transportation goals. We develop a mixed-integer programming optimization model, coded in AMPL and solved with standard software, to prescribe strategic siting and sizing decisions, and corresponding operational strategies, so as to strike a balance between equity considerations and revenue maximization. By carefully assigning weights to factors such as demand and equity, we provide valuable insights into serving diverse populations while ensuring the financial sustainability of the charging network.

WB53

Summit - 445

Capacity management under supply and demand uncertainty

Invited Session

MSOM: Supply Chain

Chair: Sammi Tang, University of Miami, Coral Gables, FL, United States

1 - Capacity Sharing Contracts and Subsidy Policies Under R&D Uncertainty

Shuyue Yi, Fudan University, Shanghai, China, People's Republic of, Xiaole Wu, Christopher Tang

Besides R&D uncertainty, firms in the semiconductor and pharmaceutical sectors face long lead times to set up production facilities. To reduce time to market, some firms make upfront capacity investment before the development of the new product is complete. To reduce capacity investment risk, capacity sharing enables a firm to recover some of its capacity investment by producing for its competitor should it fail and yet its competing firm succeed in product development. This raises several questions. Should competing firms share production capacity? Should firms establish a capacity-sharing contract before (ex-ante) or after (ex-post) R&D uncertainty is resolved? To entice firms to invest in capacity, should the government subsidize a firm's capacity investment cost or its production cost? First, we find relative to no capacity sharing, capacity sharing can entice firms to increase (decrease) their capacity when the capacity investment cost is relatively high (low). Second, ex-post capacity sharing benefits the laggard firm with a lower R&D success probability, but hurts the leading firm when it has a sufficiently high success probability. In contrast, ex-ante capacity sharing is mutually beneficial when the capacity investment cost is low. Third, without capacity sharing, capacity-based subsidy is more efficient than production-based subsidy. A caveat is that offering a higher subsidy can result in higher unmet demand (in expectation) because it entices the laggard firm to increase and the leading firm to decrease its capacity without increasing the total capacity. Fourth, under capacity sharing, capacity- and production-based subsidies are equally efficient.

2 - Optimal Reward Policies for Academic Publications

Nirmanyu Batra, The Univesity of Texas at Dallas, Richardson, TX, United States, Milind Dawande, Ganesh Janakiraman

We consider a principal and N heterogeneous agents, all of whom are risk neutral. Agents produce two goods, A and B; they exert effort over a fixed time period that results in an amount of output of A and an amount of output of B. For each unit of output that an agent produces (of A and B) the principal offers a base reward b, and for each unit of A that an agent produces the principal offers an additional reward a. The principal derives value v from each unit of A produced by an agent, and does not derive any value from units of B produced. The principal's goal is to design a reward policy that specifies base and additional reward amounts to incentivize the production of A. This problem, which we denote by P, is motivated by the prevalence of *cash-per-publication* rewards in universities, which are used to motivate faculty to publish in prestigious academic journals. When there is only a single agent (N=1), we show that it is optimal for the principal to either offer only a base reward, or only an additional reward. When there are multiple agents (N>1), we show that only-base and only-additional rewards are not optimal in general. Furthermore, we specify conditions under which it remains optimal for the principal to offer either only a base, or only an additional reward.

3 - Pricing Strategies and Capacity Constraints

Huaqing Wang, Palm Beach Atlantic University, West Palm Beach, FL, United States, Chunlin Wang, Haresh Gurnani, Saibal Ray

We consider a model in which a supplier sells a product to consumers while facing cost and demand uncertainties. We study the effect of three pricing strategies (single-pricing, demand-based, and cost-based) on supplier's profits, consumer surplus, and social welfare under unlimited capacity, limited capacity, and correlation between cost and demand.

4 - Flexibility Value of Reshoring Capacity under Policy Uncertainty

Xiao Tan, East China University of Science and Technology, Shanghai, China, People's Republic of, Panos Kouvelis, Sammi Tang

Companies that operate global supply chains are facing increasing uncertainty in the cost of imported goods—both finished products as well as raw materials. This has prompted companies to rethink the need for a diversified global supply chain, particularly by adding an onshore/nearshore production location to the current offshore location. This paper adopts a game-theoretic model to analyze a global firm's reshoring capacity, output quantity, and production decisions in the presence of domestic market competition. We account for uncertainties around market demand and import costs at both the raw-material(RM) and finished-goods(FG) level. We show that an increase in the average RM cost will reduce reshoring capacity investment, but the impact of the FG cost is ambiguous. We identify two opposite effects in the increase of FG cost: (1) an overflow demand effect—present when demand is high and exceeds an overflow threshold, and unsatisfied production must overflow to the offshore location; and (2) an output quantity effect that leads to a reduction of the reshoring investment due to a higher expected unit cost of production. The direction of change in the reshoring investment depends on the dominant effect among the above two. We also find that when the cost disadvantage of onshore relative to offshore sourcing is large (small), the presence of domestic competition can lead to more (less) reshoring. Although higher import costs hurt the global firm's profit, the domestic competitor can sometimes benefit.

WB54

Summit - 446

Food Waste

Invited Session

MSOM: Sustainable Operations

Chair: Iva Rashkova, Washington University in St Louis, St Louis, MO, United States

Co-Chair: Parisa Shahsavand, The University of Texas at Dallas, Richardson, TX, United States

1 - Reducing Food Waste in U.S. Food Banks: the Role of Distribution Strategies

Luv Sharma, University of South Carolina, Columbia, SC, United States

This study looks at different operational decisions that food banks can take to reduce food waste in their distribution operations. We utilize data from 200 food banks within the Feeding America network over a five-year horizon for this empirical investigation.

2 - The Introduction of Online Operations to Brick-and-Mortar Grocery Stores and Its Impact on Shrink

Fredrik Eng-Larsson, Stockholm University, Stockholm, Sweden, Pedro Amorim, Nicole DeHoratius, Antonio Salgado

Omnichannel grocery retailers often use stores to fulfill online orders for a variety of reasons (e.g., faster deliveries, labor availability). This strategy influences inventory management, particularly the shrink ratio (defined as the ratio of loss to sales) in two opposing ways. Fulfilling online orders has the potential to increase the sales-to-stock ratio thereby improving turnover and, consequently, the shrink ratio. However, the last-expired-first-out picking policies used by store employees to serve online customers may generate higher levels of shrink in the store. Using granular data from a grocery retailer, we study the impact that introducing online fulfillment to existing stores has on shrink. Our empirical methods include a staggered difference-in-difference approach to account for the way in which online fulfillment was adopted throughout the chain. We find the shrink ratio to increase, on average, with the introduction of online fulfillment at brick-and-mortar stores. However, substantial heterogeneity in this increase exists across product categories and stores. We exploit this heterogeneity to understand more about the dynamics of online order fulfillment and to caution retailers about several unforeseen performance impacts of online operations.

3 - Leveraging Innovative Online Platforms to Combat Food Waste

Zhangchen Hu, California State University, Fullerton, ANAHEIM, CA, United States, Wei Wei

Online platforms, such as Too Good to Go, offer an innovative approach to addressing food waste. Retailers participating in these platforms sell their surplus food in "surprise bags" at discounted prices. Considering customers' purchase decisions on such a platform and using game-theoretical models, we investigate the operational decisions of retailers and the platform. Furthermore, we analyze how factors specific to customers, the market, and the platform impact their decisions and resulting outcomes.

4 - Manufacturers' Food Expiry Dates and Waste Implications

Parisa Shahsavand, The University of Texas at Dallas, Richardson, TX, United States, Dorothee Honhon, Iva Rashkova

We study the implications of manufacturers' food expiry dates on food waste. We model a food manufacturer who sells directly to consumers and sets food expiry dates in order to maximize its revenue. Consumers shop periodically and their purchasing behavior can be either be oblivious (not paying attention to the expiry date) or aware (preference for a product with an expiry date aligned with their expectation). When consuming the product, consumers can similarly be oblivious (consumer until the product spoils) or aware (consume until the expiry date). For these four types of consumers, we compare the food waste implications as retailer and consumer waste. We further identify the extent of food wasted after expiry and before spoilage and show a non-trivial dynamic between consumer behavior and food waste.

WB55

Summit - 447

Food Waste Reduction in Supply Chains

Invited Session

MSOM: Sustainable Operations

Chair: Nicholas Petruzzi, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Aditya Vedantam, State University of New York at Buffalo, Williamsville, NY, United States

1 - The effect of store characteristics on food waste in grocery retail: An empirical analysis using Double Machine Learning

Fabian Schäfer, Technical University of Munich, Straubing, Germany, Konstantin Wink, Santiago Gallino, Alexander Hübner

Despite the well-established influence of certain characteristics of brick-and-mortar retail stores, like shelf-space and assortment planning or the size of the aisles, influencing customer behavior, little is known about the effect of various store attributes on food waste. Additionally, food waste remains a severe sustainability issue for retailers since empirical evidence about its root causes is scarce, resulting in higher carbon emissions and increased costs due to expired products. Thus, our research uses proprietary transaction data and geo data from a European retail chain partner company to investigate the relationship between store-specific characteristics and in-store waste levels. The data covers 315 retail stores in urban and rural areas. By applying the Double Machine Learning algorithm XGBoost for causal inference, our findings indicate that store-specific characteristics, such as grocery store density, delivery time, and delivery frequency, significantly impact food waste levels relative to its revenues. Our study fills a gap in the literature on food waste in grocery retail by providing empirical evidence on the store characteristics that lead to higher food waste levels. Furthermore, this study creates awareness and offers novel managerial insights for practitioners considering food waste when planning strategic store locations.

2 - Information Disclosure for Perishables: Profit and Food Waste Implications

Fan Zhou, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Ekaterina Astashkina, Ravi Anupindi

The freshness of food, as measured by its residual shelf life (RSL), is often variable and unseen by consumers in online grocery retailing. This study examines the impact of disclosing RSL information to consumers on profit and waste. To that end, we consider two information schemes: a visible freshness scheme, under which consumers observe the food's RSL sold by the retailer; and an invisible freshness scheme, under which RSL is unobservable. We build a discrete-time game theoretical model with two players: the retailer and the consumer. The retailer announces the food's price under a certain information scheme, and the consumer, based on the available information, decides on the purchase quantity that maximizes the expected utility. The utility calculation takes into account household waste, which is jointly influenced by the purchase quantity, the RSL of the food at the time of purchase, and the food type. We find that the visible freshness scheme results in higher profits compared to the invisible freshness scheme due to: (i) the presence of pricing flexibility, which leads to higher sales; and (ii) consumers' risk aversion towards uncertain freshness. However, disclosing RSL to consumers does not always result in lower waste and, in fact, depends on the food type. In particular, for food that is less prone to waste, disclosing the RSL actually increases household food waste. Conversely, for food that is more prone to waste, disclosing RSL delivers a win-win by boosting the profit and reducing waste simultaneously.

3 - Strategic Sell-By Dates: Implications for Retail Food Waste

Karthik Murali, Oregon State University, Corvallis, OR, United States, Aditya Vedantam, Nicholas Petruzzi

Grocery stores in the U.S. throw away 16 billion tons of food each year due to expiration. In this study, we develop a game-theoretic model to examine the strategic motivations behind a manufacturer's choice of sell-by dates and the corresponding implications for food waste at the retail level.

WB56

Summit - 448

Optimizing Aviation and Air Mobility

Contributed Session

Chair: Randy Grivel, Arizona State University, Tempe, AZ, United States

1 - Integrated Recovery of Passenger and CargoConsidering the Belly Space of Passenger Aircraft

Sirui Wang, Harbin Engineering University, Harbin, China, People's Republic of, Yuzhen Hu

Thanks to the recovery of the international passenger market this year, the cargo capacity of passenger aircraft has increased significantly. It is necessary to rearrange cargo itineraries and jointly consider the coordinated recovery of passenger itineraries under the situation of cargo carried in the belly hold of passenger aircraft. We establish an integrated passenger-cargo recovery model and apply time-space network models to depict the transfer networks of aircraft, passengers, and cargo respectively. Taking into account the characteristics of cargo transportation, such as no restriction on the number of transfers under the premise of coordinating costs, we propose three strategies for the cargo transfer network to improve computational efficiency. Then, we conduct a case study using real data from a large Chinese airline and simulated data on cargo transportation. By comparing with the currently used model of airline, we explore the solving performance of the two recovery models. It is found that an increase in the number of cargo transfers leads to a situation where "the solution results are better but the solving efficiency is reduced," thereby validating the two improvement strategies proposed in this paper. And the first strategy can effectively improve the solving efficiency of the integrated model, increase the generation time of the transfer network, but overall reduce the solving time; the second strategy can significantly improve the solving efficiency at the cost of a deviation of less than 5% from the target value; and the third strategy can largely balance solving results and computational efficiency.

2 - Optimization-Based Simulation of Target Time Over For Flights Within the Decentralized ASEAN Region

Darryl Teo, Singapore University of Technology and Design, Singapore, Singapore, Peter Jackson, daniel delahaye

Air Traffic Flow Management (ATFM) tactical measures such as the imposition of acceptance rates and calculated take-off times require information sharing and collaboration between Air Traffic Control sectors. In a decentralized system such as Southeast Asia this is a challenge. To evaluate the potential benefits of different measures under different information sharing regimes, it is useful to construct a rolling horizon simulation of the decentralized system. An essential component of such a simulation is a model of each Flight Information Region (FIR) as a separate decision-making entity. We formulate the FIR decision problem as an optimization problem, the Waypoint Flow Regulation Problem (WFRP), which is to be solved for each FIR at each time step of the simulation using the local information available to each FIR at that time. We propose two methods to solve the WFRP: Gradient Descent Ascent (GDA), and Simulated Annealing (SA). Both heuristics were applied within the integrated simulator to published flight schedules, under two capacity scenarios, for the Southeast Asia region. Both algorithms achieved near-feasible solutions. SA yielded better flight delay performance but GDA was an order of magnitude faster.

3 - HEAT: Hub Efficiency Analytics Tool

Majid Akhgar Farsani, American Airlines, Grapevine, TX, United States, Billy Wang, Duaa Serhan, SHREEJIT BANDYOPADHYAY

In this work, we will present a large-scale optimization model developed by the Operations Research and Advanced Analytics team at American Airlines that aims to optimize the flight delay and cancel decisions as potential responses to operational capacity reductions at the airline's major hub airports. Adverse weather conditions, such as severe thunderstorms in the summer, can pose challenges to the safe operation of flights, resulting in capacity reduction on the arrival side, or the departure side, or both. Historically, the airlines have two options to manage this critical operational problem: either canceling flights proactively or collaborating with the FAA to implement air traffic management initiatives. However, both options tend to have their respective drawbacks. The model we designed is to address some of these drawbacks by incorporating a rich dataset of airline internal data and optimizing the delay and cancel decision jointly. Since its initial deployment in April 2022, this new model has been used more than 100 times at the airline's largest hub airports and prevented thousands of flight cancellations across the network. The use of the model has contributed to American Airlines' record operational performance in 2023.

4 - Aircraft Line Maintenance Scheduling: Q-Learning Approach

Syed Shaukat, University of New South Wales, Sydney, Australia

Recent advances in optimal airline scheduling have seen a shift towards data-driven and AI-based learning approaches, indicating great potential to complement or even replace conventional optimization-based models. Conventional optimization methods encounter significant challenges when applied to real-world conditions, including the uncertainties and nonlinearities inherent in the airline environment. This paper delves into the application of Q-Learning, a prominent reinforcement learning (RL) method, within the dynamic and complex realm of aircraft line maintenance job scheduling. Employing the reward-centric properties of reinforcement learning, the proposed methodology adeptly navigates the operational intricacies of airline scheduling.

The results demonstrate that the Q-Learning approach is capable of generating solutions comparable to those of conventional optimization models when tested with industry-inspired datasets. This AI learning method adeptly balances maximizing aircraft availability with maintaining stringent maintenance schedules, potentially affording airlines the operational agility necessary for superior decision-making, cost savings, and increased flexibility. By advocating a standardized and uniform approach, this AI-based methodology paves the way for integrated solutions to aircraft maintenance job scheduling, aircraft maintenance check scheduling, and aircraft tail assignment problems.

5 - Scheduling Resources in an Airport Security Screening Process with Break, Overtime, and Skill Considerations

Randy Grivel, Arizona State University, Tempe, AZ, United States, Ron Askin, Jorge Sefair, Jian Jiao

Airport security services in the United States can be modeled using sequential heterogeneous servers with dynamic processing rates and pooled resources. We model this system as a mixed integer program with the objective of minimizing the combined expected cost of passengers' wait time and operational cost of staff overtime and idle time. Our model includes realistic operational rules such as the ramp up and down processing rates when opening and closing stations, staff movement across facilities, working shifts, skill set of officers, and design of breaks and overtime plans. A novel method is used for determining staffing requirements of a baggage screening area. We illustrate the performance of our models using a real-based instance from an US airport.

WB57

Summit - Terrace Suite 1

Incorporating AI into Healthcare Delivery

Invited Session

Health Applications Society

Chair: Sidian Lin, Harvard University, Cambridge, MA, United States

1 - Exploring the Effects of Retrospective Study Design on Non-Alcoholic Fatty Liver Disease Prediction

Mary Ogidigben, Pennsylvania State University, State College, PA, United States, Paul Griffin, Soundar Kumara

Non-alcoholic fatty liver disease (NAFLD) affects 25% of the U.S. adult population and is the most common chronic liver disease. The condition is asymptomatic in most individuals. Progression of the disease leads to liver cirrhosis and is a leading indication for liver transplant. Although universal screening is not recommended, early detection is beneficial since lifestyle changes can control or reverse the condition. Several NAFLD risk prediction models have been developed and often modeled with machine learning techniques. However, most previous methods have been developed for homogenous populations and there is no consensus on the best predictive factors to use. We explore how retrospective study design affects feature selection and accuracy to construct a framework researchers can use to develop fair algorithms for early chronic disease diagnosis across a variety of patient classes.

2 - Development and Validation of a Reinforcement Learning Algorithm for Personalized Treatment in Bipolar Disorder

Sidian Lin, Harvard University, Cambridge, MA, United States, Soroush Saghafian, Jessica Lipschitz, Katherine Burdick

This paper presents a novel application of offline Multi-Agent Reinforcement Learning (MARL) for optimizing bipolar disorder (BD) treatments. Using Fitbit data and biweekly self-reports from BD patients, we develop an approach integrating the Implicit Constrained Q-Learning (ICQ) algorithm with copula-based dependency modeling. The proposed MARL demonstrates robust performance in optimizing individualized treatment plans, reducing mood episode frequency and potential suicide rates.

3 - Sequential Decision Making with Sparsely Observed Rewards

Daiqi Gao, Harvard University, Cambridge, MA, United States, Hsin-Yu Lai, Susan Murphy

Cardiac rehabilitation (CR) is an outpatient risk-reduction program for patients with cardiovascular disease, involving supervised exercise, dietary counseling, and stress management. However, sustaining the heart-healthy lifestyle changes adopted during CR after patients return to the obligations of everyday life remains a significant challenge. Our aim is to improve users' commitment to physical activity (PA) immediately after they leave the CR program by providing effective ongoing support through Mobile health (mHealth). This will be achieved through notifications on mobile devices prompting short bouts of activity. However, information about commitment to PA is collected through weekly or monthly surveys, resulting in sparsely observed rewards. We employ reinforcement learning to determine when and whether to deliver notifications. Causal information provided by domain experts is leveraged to speed up learning. We will also discuss practical challenges in designing the learning algorithm and constructing a simulation testbed.

4 - Enhancing Fairness and Accuracy in Mammography Screening Using
a Custom Ensemble

Mehmet Ahsen, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Mehmet Ayvaci, cetin ozeken

As AI algorithms become increasingly prevalent in healthcare, ensuring both accuracy and fairness in algorithmic decision-making poses a dual challenge for healthcare organizations. We tackle this challenge by leveraging ensemble methods to promote fairness while maintaining accuracy, specifically in the context of mammography screening. This setting is particularly vulnerable to disparities in algorithmic bias among diverse racial groups. Moreover, multiple commercially available AI algorithms for mammography demonstrate varying accuracy across distinct subpopulations. We first develop a statistical analytical model to establish conditions for constructing a fair ensemble algorithm, even when individual algorithms exhibit unfairness. Through structural analysis, we demonstrate the trade-off between fairness and performance optimization by comparing the fair ensemble with the optimal ensemble. Then, we demonstrate our approach using twelve algorithms from
the Digital Mammography Dialogue on Reverse Engineering Assessment and Methods (DREAM) Challenge, trained and validated on mammograms from Kaiser Permanente Washington (KPW). We applied those algorithms to a large, geographically diverse screening population at a different healthcare organization in West Coast. The fair ensemble aims to achieve equitable performance in, contributing to more equitable healthcare outcomes in breast cancer screening. Our empirical analysis showcases the effectiveness of ensemble methods in improving both generalization and fairness in mammography screening. In doing so, our approach considers practical implementation factors, leveraging algorithm diversity and appropriate weighting to optimize performance while promoting equity.

WB58

Summit - Terrace Suite 2

Advances in healthcare: Methods and applications

Invited Session

Health Applications Society

Chair: Mark Van Oyen, Industrial & Operations Engr., Ann Arbor, MI, 48109, United States

Co-Chair: Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, 48105, United States

1 - Ai Assistant for Highly Personalized and Dynamic Causal Decision-Making in Cancer Treatments

sina Aghaei, Harvard University, Cambridge, MA, United States, Jacob Jameson, Giuseppe Tarantino, David Liu, Soroush Saghafian

Large language models (LLMs) have demonstrated impressive capabilities in various domains, including healthcare. Existing work on LLMs in the medical field has mainly evaluated general-purpose models on tasks like information retrieval, summarization, question answering, and diagnostics. Less effort has focused on enhancing LLMs for impactful clinical uses such as those that require complex causal reasoning. This paper bridges this gap by developing an AI system tailored to optimizing treatment regime recommendations and assisting providers in their complex causal reasoning and decision-making processes. In close collaboration with the Dana-Farber Cancer Institute, we harness detailed data on patients with melanoma, spanning clinical records, diagnostic imaging, pathology reports and gene sequencing information, among others. We develop a multi-modal LLM and fine-tune it to have causal reasoning capabilities using novel causal reinforcement learning techniques. This enables the AI system to provide highly personalized and dynamic (i.e., fully adaptive) treatment recommendations via causal reasoning.

2 - A Robust Approach to Locating Vaccination Centers During A Pandemic: A Problem Driven by Herd Immunity and Decision-Dependent Concave Utility

Shibshankar Dey, Northwestern University, Evanston, IL, IL, United States, Esma Gel, Sanjay Mehrotra

This study introduces an optimization model for determining the location of vaccination centers (VC) while accounting for distributionally-robust decision-dependent utility, herd immunity, equity and facilities' end-of-life value. We consider the herd-immunity function as a potentially useful surrogate to the epidemiological compartmental model. We explore both linear and concave utility functions which, besides evaluating the utility of VC location, also have the potential to be a measure of vaccine hesitancy. Addressing uncertainty in the parameters defining utility and the herd immunity function, we first reformulate the problems respectively as two-stage stochastic mixed-integer bilinear and mixed-integer biconvex programs. Considering the inability of GUROBI to handle resultant bi-convex problem, we develop a decomposition linearize-branch-union approach with an embedded exploration-based heuristic. Our proposed decomposition algorithm iteratively adds optimality cuts in x-space (first-stage variable space) by linearizing, branching, union of the leaf nodes of branch-and-bound tree and finally projecting the union set onto x-space. Novelty lies not only in ensuring global validity of those cuts but also the adaptive linearization scheme used for bi-convex problem linearization. Using the proposed solution framework, we then solve several instances and interpret how location decision changes with anticipated herd immunity levels, equity thresholds, robustness considerations. Furthermore, we examine how concave utility affects the location decision and increases vaccination level in comparison to linear utility and no utility at all. We finally discuss the potential and flexibility of our proposed decision-dependent utility-based model and algorithm to solve diverse assignment problems that dominantly arise in the healthcare domain.

3 - Applications of Non-Stationary Bandits in Personalized Healthcare

Yonatan Mintz, University of Wisconsin Madison, Madison, WI, United States, Katherine Adams, Qinyang He

Mobile interventions have become commonplace in many healthcare applications ranging from diabetes management, mental health, addiction management, and weight loss. With this level of data availability interventionists and clinicians are now able to provide personally tailored interventions to individual participants to improve overall efficacy and adherence. A key aspect of these treatments is that the decision makers must use online intervention information to both learn the effect of various treatments on individual participants as well as decide on the most effective treatment options. This setting can thus be modeled through the use of multi-armed bandit (MAB) models which are designed to effectively trade off exploration (learning the parameters of the system) and exploitation (finding the optimal treatment choice). In contrast to classical applications of MAB models where only a single action or time invariant set of actions may be optimal, in the context of personalized interventions data is often non-stationary requiring MAB approaches that can account for frequently changing optimal actions. Moreover, interventionists are often faced with resource constraints such as limited numbers of nutritionists for counseling in the case of weight loss interventions, or limited financial budgets for contingency management interventions in addiction management. In this talk, I will discuss some of the recent advances made by my group in adapting non-stationary MAB models to these resource constrained settings. I will discuss both new theoretical guarantees that we can obtain for such algorithms and discuss several computational case studies showing their practical performance.

4 - Patient Management in the Neuro-ICU: to Scan OR Not to Scan?

Agni Orfanoudaki, Oxford University, Oxford, United Kingdom, Panos Tsimpos, Ethan Phillips, Charlene Ong, Stelios Smirnakis

Stroke survivors in the neuro-intensive care unit are at an increased risk of rapid cerebral edema development. Diagnosis requires cranial radiology imaging. However, high operational costs, radiation exposure, and scanner scarcity render hyper-aggressive scanning detrimental. To address this problem, we propose a partially observable Markov decision process approach with a one-step memory that allows us to dynamically capture the patient trajectory and suggest medical screening at the onset of clinical deterioration. We analytically characterize the derived optimal policy and identify a critical threshold for medical intervention. Moreover, we demonstrate how supervised machine learning models could improve the proposed policy by estimating a personalized patient kernel. We validate our framework leveraging data from two academic medical centres, highlighting the value that our approach could bring to real-world settings.

5 - The Costs of Inaccurate Resupply Information During Drug Shortages

Jacqueline Griffin, Northeastern University, Boston, MA, United States, Noah Chicoine

During drug shortages, pharmaceutical manufacturers may communicate estimated release dates (ERDs), communicating when the next batch of a specific product should enter the supply chain. Hospital pharmacists use ERD information to inform their inventory management and sourcing decisions. However, ERD information tends to be inaccurate, which has the potential to cost hospitals additional time, labor, and money if they make decisions that fully rely on this information. We demonstrate this phenomenon using a stylized model of optimal drug sourcing strategies using ERD information. We show that inaccurate ERD information can cause hospitals to invoke costly last-minute practice changes. We also demonstrate that more explicit descriptions of uncertainty ERDs could reduce these costs. This work highlights a possible area for improvement in increasing the resiliency of pharmaceutical supply chains and improving patient care at hospitals.

WB59

Summit - Ballroom 1

Planning and Operations of Power Systems

Contributed Session

Chair: Hanieh Rastegar, University of Illinois at Chicago, 821 W Bradley Pl, Chicago, IL, 60613, United States

1 - Capturing the Role of Peaking Power Plants in a Decarbonized Electricity System

Wesley Cole, National Renewable Energy Laboratory, GOLDEN, CO, United States, Brian Sergi

Electricity systems rely on peaking power plants to maintain system resource adequacy. These plants typically operate for only a small number of hours each year when dispatchable generation needs are greatest. Modeling of decarbonized electricity systems shows a strong role for peaking power plants such as hydrogen fuel cells, hydrogen combustion turbines, or uncontrolled natural gas plants with carbon offsets. However, the representation of these power plants can have considerable impact on the specific role that these types of plants play. This work will showcase several approaches for capturing peaker plant behavior in a national-scale capacity expansion model of the U.S. electricity system. It will highlight how resource adequacy choices, power plant flexibility choices, and efficiency choices impact model outcomes. For example, the choice between including a simple-cycle hydrogen combustion turbine (which has a lower efficiency) and a combined-cycle hydrogen combustion system (which has a higher efficiency) can have substantial impacts on the cost of achieving a fully decarbonized electricity system. This work will highlight the importance of peaker plant design decisions for electricity system planning models that are being used to study decarbonization scenarios.

2 - Modeling peak electricity demands across distribution grid types

Christine Gschwendtner, Massachusetts Institute of Technology, Cambridge, MA, United States, Seth Blum, Lauren Aguilar, Jessika Trancik

The electrification of heating and transport can cause increased peaks in electricity demand, if left unmanaged. These peaks may strain electric grids and require a costly expansion of the electricity supply. The combined integration of heat pumps and electric vehicles, and potential compound peak effects, may be especially challenging and should be investigated. These effects can be particularly strong at small spatial scales, i.e., for distribution grids, requiring infrastructure upgrades. Here we investigate these potential changes in peak electricity demand in different distribution grid types considering various adoption scenarios for heat pumps and electric vehicles. We characterize distribution grid types in the US according to characteristics that include the share of residential and commercial buildings, urban and rural areas, and climate zones. We simulate hourly electricity demand profiles with a focus on the temporal and spatial variation of weather events over a multi-year timescale. Analyzing the magnitude, duration, frequency, regularity, and timing of the resulting peaks for each technology, and across residential and commercial applications, allows us to identify mitigation strategies for peak reduction, such as demand response. These insights can support the development of demand-side management schemes and can inform grid investment decisions.

3 - Effects of Storage Participation on Wholesale Electricity Markets

Zhenhua Zhang, University of California, San Diego, Mountain View, CA, United States, Michael Davidson

The strategic behaviors of storage systems that participate in wholesale electricity markets can influence market prices, system costs, and profit distributions among generation types. We develop a bi-level model to examine how these impacts vary across participation requirements in the Western Electricity Coordinating Council region. The upper level includes a storage profit maximization model and the lower level represents the market operator's economic dispatch problem. We find that strategic behaviors can significantly increase storage profit but increase system cost only by a small margin under various renewable penetration levels relative to the case when storage is ISO-controlled. Market prices decrease during charging events but remain similar during discharging events, subject to electricity demand conditions and discharge offer prices. For storage that also participates in resource adequacy programs with minimum state-of-charge requirements, a low non-compliance penalty can help recover storage investments without significantly increasing system costs. Future designs on wholesale participation models and resource adequacy programs should consider how ISO requirements and storage strategic behaviors interact and evaluate the trade-offs between generator profitability and system costs.

4 - Post Disaster Restoration of Electrical Distribution Networks: Introducing a new variant of the TSP

Ran Wei, University of Washington, Seattle, WA, United States, Arindam Das, Daniel S. Kirschen, Payman Arabshahi

The costs of severe weather events, both tangible and intangible, on the overall power grid (generation, transmission, and distribution systems) can be exorbitant, depending on the scale of the event. Optimal restoration of distribution networks following a major weather event is a multifaceted optimization problem which involves pre-event as well as post-event decision making under uncertainty. We concentrate on post-event activities, specifically, optimal repair scheduling such that the distribution network can be brought back online as quickly as possible, thereby minimizing the human, social, and economic costs after a major event. Since a major limitation to the restoration process is the lack of timely and accurate information about the scale and locations of damages, we adopt a rolling horizon restoration framework which is naturally conducive to periodic resequencing of restoration activities, if necessary, due to repair time overruns and/or uncertainty in damage information. Considering both repair times and travel times of crews, we show that the optimal restoration problem can be formulated as a combinatorial optimization problem which closely resembles an instance of a cost-constrained reward-maximizing multiple traveling salesman problem (TSP) on doubly weighted graphs (weights on both nodes and edges, modeling repair times and travel times respectively). This is a new variant of the TSP and can be viewed as a generalization of the Selective TSP. We propose a mixed integer linear programming model to solve the above problem and validate it on standard IEEE test distribution networks as well as randomly generated network instances.

5 - A Two-Stage Stochastic Model for Mixed-Fleet Pickup and Delivery Routing Problem with Temperature Considerations

Hanieh Rastegar, University of Illinois Chicago, Chicago, IL, United States, Amir Shafiee, Xi Cheng, Jane Lin

Transportation is a significant source of CO_2 emissions, with freight operations being especially impactful. This study explores sustainable alternatives in freight transportation by integrating Electric Vehicles (EVs), known for their zero tailpipe CO_2 emissions. The adoption of EVs in distribution networks poses challenges such as high initial costs, restricted range and payload capacities, and long recharging durations. To address these issues, we introduce the Pickup and Delivery Multi-Fleet Vehicle Routing Problem (PDMFVRP). This model aims to minimize total costs, encompassing vehicle acquisition, travel time, and energy usage, while also incorporating the effects of temperature on EV performance. We employ a two-stage stochastic programming approach to optimize the composition of diesel and electric fleets, considering the variable impact of temperature on EV efficiency and routing. This approach aims to minimize both total operational costs and CO_2 emissions. Utilizing the epsilon constraint method for multi-objective optimization, we seek an optimal trade-off between cost efficiency and emission reductions, factoring in the complexities of EV charging demands and constrained driver hours. Our analysis indicates that although the shift to EVs may lead to higher costs and increased vehicle miles due to frequent charging stops, it substantially lowers tailpipe emissions. This study presents a detailed and sustainable strategy for the modernization of goods distribution, aimed at reducing environmental impacts.

WB60

Summit - Ballroom 2

Optimizing Electric Vehicle Charging Infrastructure

Contributed Session

Chair: Jiangxue Han, N/A

1 - Enhancing electric vehicle charging demand prediction with graph neural networks

Yuanjie (Tukey) Tu, University of Washington, Seattle, WA, United States, Mayuree Binjolkar

As Electric Vehicles (EVs) become increasingly popular, predicting EV charging demand at public EV charging stations has been critical to help optimize operations and prevent grid overloading. In this study, we leverage graph neural network (GNN) to predict EV charging demand, incorporating both multivariate historical EV charging data and spatial-temporal interactions between EV charging stations. Our approach is compared with other statistic and deep learning models including autoregressive moving average (ARMA), multiple linear perceptrons (MLP), convolutional neural networks (CNNs), and long short term memory (LSTM). Moreover, we investigated the model explainability of GNN. As a result, our approach is conceptually advantageous to other methods, which do not consider the interactions between EV charging stations or model explainability. Empirical results on a time-series EV charging dataset of 47 charging stations show that CNN outperformed start-of-art methods. These findings provide actionable insights for city planners and engineers to anticipate peak usage periods, optimize operations, and help with power grid stability and sustainability.

2 - Optimal Placement of Electric Vehicle Charging Stations

Iason Liagkas, University of Michigan Ann Arbor, Ann Arbor, MI, United States, Neda Masoud

As the demand for electric vehicles (EVs) continues to surge, efficiently integrating EV charging stations within urban settings emerges as a critical challenge. This paper investigates the strategic placement of these stations to achieve three primary goals: reducing overall traffic congestion, minimizing government expenditure on infrastructure, and lowering the societal costs associated with conventional internal combustion engine vehicles. Our approach considers the availability of public transportation as an alternative to EV usage, ensuring a comprehensive analysis of urban mobility. Furthermore, Benders decomposition is adopted as a robust solution strategy, with its efficacy illustrated through detailed numerical examples. This study not only provides practical insights for urban planners but also contributes to the broader goal of making cities more sustainable and efficient in the face of rising EV adoption.

3 - Optimum Configuration of Fast Charging with Integrated Distributed Energy Resources during Fluctuating Demand and Extreme Weather Conditions

Hamid Mozafari, Michigan State University, East Lansing, MI, United States, Kunle Adeyemo, Chenyang Deng, Mehrnaz Ghamami, Annick Anctil, Ali Zockaie

The rapid transition towards electric vehicle (EV) necessitates a well-established network of direct current fast charging (DCFC) stations and a robust power distribution system. Recent battery technologies have mitigated range anxiety for urban EV users, though challenges persist for long-distance travel. This study aims to determine the optimum charging infrastructure for EVs incorporating second life batteries (SLBs) and photovoltaics (PVs) as distributed energy resources (DERs) under fluctuating travel demands and temperatures. The proposed framework consists of two optimization models: the first minimizes the total system cost, including investment (i.e., charger cost) and user (i.e., queueing and detour delay) costs, while the second, executed by Homer Pro, determines the optimal capacity of DERs at each charging station. DERs not only reduces the system costs and alleviates the strain on the electricity grid network during peak-demand seasons and hours but also DERs also mitigates the environmental impact by repurposing used EV batteries. The results demonstrate that incorporating SLBs and PVs into DCFC stations is not only cost-effective but also significantly alleviates the grid's load, besides the environmental benefits of reusing the EV batteries. Integrating the optimal capacity of SLBs and PVs into DCFC stations has the potential to diminish the required grid power by up to 70%, and decrease the total system costs by 59%, contingent upon the DERs acquisition scenarios and the charging schedule approach.

4 - A Pricing Model for Enhancing the Utilization of Public Electric Vehicle Facilities: A Case Study in Hong Kong

Jiangxue Han, The University of Hong Kong, Hong Kong, Hong Kong, Yong-Hong Kuo

With the wide acceptance of electric vehicles (EVs) and the significant increase in the number of EVs, greater attention has been paid to the development of public EV charging facilities. However, issues such as constrained charging spaces, EV overcharging, and inconvenient charging locations, might lead to difficulties in achieving a high utilization of public EV charging facilities. Existing research primarily focuses on enhancing the spatial planning of EV charging stations. In this research, we provide a pricing perspective as a more flexible and adaptive complement to optimize the utilization of the established EV charging infrastructure.

We focus on the case of Hong Kong in this work. While the Government has adopted an (almost) uniform pricing policy, we develop a more effective pricing scheme for the public EV charging services in the city. Based on the primary factors that may influence drivers' choice in EV charging facilities, we build a linear optimization model to explore the mutual influences between these factors, with the objectives to improve the utilization of public EV chargers and to satisfy the EV charging demands. Numerical experiments are conducted based on real data to examine the performance of the proposed pricing mechanism.

WB61

Summit - Ballroom 3

Application of Machine Learning Methods

Contributed Session

Chair: ekrem duman, Ozyegin University, Istanbul, Turkey

1 - Utilizing Large Language Models for Improved Missing Data Handling in Classification Tasks

Francis Kim, Sungkyunkwan University, Suwon, Korea, Republic of, Dohyun Bu, Chanho Kim, Jong-Seok Lee

In this research, we propose a novel method for handling missing values in tabular data by utilizing Large Language Models (LLMs) for classification tasks. Traditional approaches to handling missing values, such as imputation or deletion, often introduce bias and decrease prediction accuracy. Our method addresses these issues by systematically augmenting the complete training dataset with various combinations of missing values and converting each instance into a textual format. This enables the Large Language Model (LLM) to learn from the augmented data during fine-tuning. The process includes three main steps: data augmentation through the introduction of missing values, converting these augmented instances into text, and fine-tuning the LLM using this textual dataset. The LLM is also trained to predict the target class based on the available information, thereby effectively learning the marginal probabilities of the target variable given the observed data. Experimental results demonstrate the effectiveness of our method across different missing data mechanisms (MCAR, MAR, MNAR), showing improved performance over traditional imputation methods. Our approach ensures robust predictions and maintains high accuracy, even in situations where the data contains a significant number of missing values, making it a valuable tool for domains frequently encountering incomplete datasets.

2 - Better Generative AI, Better Prompts: The Ongoing Need for Prompt Engineering

Eaman Jahani, University of Maryland, Washington, DC, United States, Hong-Yi TuYe, Joe Zhang, Mohammed Alsobay, Benjamin Manning, Siddharth Suri, David Holtz

The extent to which generative AI can provide value in any context is highly contingent on the quality and efficacy of the textual instructions, “prompts”, that users provide to the model. The practice of creating prompts for generative AI has come to be known as “prompt engineering” and some firms have begun creating dedicated prompt engineer job positions. Although there are already a number of studies documenting the quality gains due the advances in generative AI models, there has been very little focus on how more powerful models influence the necessity for prompt engineering. Several scholars have argued that as generative AI models become more proficient in understanding human language, the need for meticulously crafted prompts will diminish. We investigate how advances in generative AI models interact with the importance of prompt engineering and show that at the very least over the short term, prompting techniques remain important in generating higher quality content as models become stronger. We show that the transition for DALL-E2 to DALL-E3 leads to significant improvements in the quality of user-requested images. We further decompose this gain to two sources: improvements directly resulting from the model enhancement and, improvements due to changes in prompting techniques indirectly induced by the stronger model. Each of the mediating sources contribute to about half of the overall performance gain. Our results suggest that as generative AI technology advances, both the inherent capacity of the models and the more sophisticated prompting strategies they enable are critical for generating high quality content.

3 - Generative AI in Recruitment: A Comparative Study of Llama2 and ChatGPT for Resume Screening

Wenyue Xi, New York University, New York, NY, United States, Oscar Wan, João Sedoc

Generative AI has the potential to enhance hiring efficiency at scale; however, the integration of Large Language Models (LLMs) into such processes requires examination of trade-offs. Thus, we study a popular open-source model, Llama 2-13B, and a close-source model, ChatGPT, in the context of the recruitment process. We focus on their lexical processing, explanation, and decision-making abilities. We aim to address critical concerns, such as enhancing the financial accessibility of open-source LLMs, addressing privacy concerns, improving explainability in hiring processes, and ensuring transparency and reproducibility in their outcomes. We use a resume dataset (Jiechieu & Tsopze, 2020) and a job description dataset from PromptCloud and DataStock. First, we assess the capability of Llama2 and ChatGPT to convert resumes into structured data and effectively align them with job descriptions. Next, we utilize methodologies such as word embedding, BERTopic, and LlamaIndex to determine how well these models match resumes with job descriptions based on their suitability. Additionally, we investigate the key input features that influence decision-making in Llama2 during the resume screening process and employ Integrated Gradients and Chain-of-Thought Prompting to explore the models’ explainability (Parasurama et al., 2022). This study emphasizes the necessity of evaluating Generative AI applications in recruitment to enhance efficiency and ensure fairness in the automated hiring process.

4 - Development of Collection and Churn Scorecards

ekrem duman, Ozyegin University, Istanbul, Turkey

This article addresses the challenges faced by banks in managing non-performing loans (NPLs) and optimizing collection strategies for delinquent customers. The study presents a systematic approach to optimize collection strategies, aiming to maximize the total probability of payment before litigation for each customer while minimizing the risk of customer churn. The study proposes the development of collection and churn scorecards framework to predict the effectiveness of various collection actions and their potential impact on customer churn. Using logistic regression models, the author analyze historical data to identify key predictors of payment and churn probabilities, incorporating action severity and capacity constraints into the modeling process.

WB62

Summit - Signature Room

Platform Design and Operations

Invited Session

Auctions and Market Design

Chair: Yunke Mai, University of Kentucky, Lexington, KY, United States

1 - Optimal incentive design for ongoing book writers with freemium-to-premium reading mode

Lang Fang, Dongbei University of Finance and Economics, Dalian, China, People's Republic of

The rise of online paid reading has resulted in an increase in publishing platforms offering ongoing books for sale online. Instead of releasing complete books, writers release several chapters for free and then utilize a pay-per-chapter option to continue releasing additional chapters based on reader engagement (freemium-to-premium reading mode). Writers adjust chapter quality based on reader feedback in this process. For platforms, how to set the chapter price and the profit sharing ratio of writers is a key decision. Our research introduces two principal-agent models to assist platforms in effectively implementing (i) myopic and (ii) forward-looking chapter pricing and profit sharing ratio-based contracts to maximize platform benefits. For the myopic contract design, which means that the platform only focuses on maximizing profit for one chapter. We develop a principal-agent model to determine the price of chapters and the profit sharing ratio of writers. The optimal solution is provided by model simplification and specific search algorithm respectively. For the forward-looking problem that the platform aims to maximize the total profit of an ongoing book, we develop an efficient solution procedure to simultaneously set the chapter price and decide the profit sharing ratio of writers over a planning horizon. Additionally, we also explore the impact of the number of free chapters on platform revenue optimization. We demonstrate the superiority of our solution strategies for both myopic and forward-looking problems against multiple benchmark methods used in practice. Finally, we present several valuable insights into writer and ongoing book management for these platforms.

2 - A Hybrid Solution to the Print-on-Demand Platforms: Can Less be More?

Pengyu Wang, University of Washington, Foster School of Business, Seattle, WA, United States, Yuankun Li, Masha Shunko, Yong-Pin Zhou

Print-on-Demand (POD) platforms are emerging as popular online marketplaces where individual content creators can upload unique graphic designs for customers to select, which are then printed on various blank products and shipped to the customers by the platform. The platform incurs all the costs and shares the revenue with the content creators based on a pre-determined schedule. Currently, POD platforms use a make-to-order (MTO) production model, which provides flexibility and minimizes inventory cost, but comes at a high production cost. We propose a novel hybrid production strategy that supplements the MTO model with the cost-effective make-to-stock (MTS) model. To make sure that the new system is incentive compatible, we propose three different implementation strategies that encourage content creators to focus on increasing demand for each design rather than flooding the market with excessively many designs. Our solution leads to substantial reductions in unit production costs and significant profit increases for both the platform and the content creators.

WB63

Regency - 601

The Future of AI

Invited Session

Information Systems

Chair: Chen Liang, University of Connecticut, Storrs, CT, United States

Co-Chair: Yinliang Tan, University of Houston, Houston, TX, United States

1 - Would AI Replace Friends? Examining the Impact of Ai Interaction Frequency on Social Relationships

Zenan Chen, The University of Texas at Dallas, Richardson, TX, United States, Teng Ye, Jason Chan

This study examines the potential of large language model (LLM)-based artificial intelligence (AI) companions to fulfill human socialization needs, an area of increasing interest as these digital entities become more integrated into daily life. With their ability to understand and respond in a human-like, empathetic manner, these AI companions have quickly attracted millions of users. This research aims to understand whether interactions with LLM-based chatbots could replace human relationships, a question with significant implications for our comprehension of digital companionship and the socio-psychological impacts of technology. Such insights are crucial for informing future public policy on LLM regulation and the design of socially responsible AI technologies.

To explore this, we design a randomized experiment using a cross-platform AI companion powered by a large language model featuring customizable 3D avatars. The study focuses on measuring changes in the depth and breadth of participants' social interactions with friends and family via social media messaging, before and after the introduction of the LLM-based companion. This ongoing work promises to shed light on the evolving dynamics of human-digital relationships and their broader societal and theoretical implications.

2 - Beyond Automation: the Impact of Robotaxi Services on Ride-Hailing Consumer Behavior

Chen Liang, University of Connecticut, Storrs, CT, United States, Tong Shen, Xinxin Li, Chunxiao Li, Shuliu Yuan

Robotaxi technology, an innovative development within ride-hailing services, is undergoing regional deployments across different countries, yet its influence on consumer demand remains largely unexamined. Our study employs a difference-in-differences (DiD) model to estimate the causal impact of robotaxi usage on consumers' ride-hailing demand by comparing the behavior or consumers experienced robotaxis services with those who are did not. Our identification strategy leveraging a unique identification strategy based on the algorithmic, quasi-random distribution of robotaxi services to consumers. Our identification strategy exploits the algorithm-driven, quasi-random allocation of robotaxi services to consumers when they requested robotaxi services, wherein some consumers experience them while others do not. Our findings reveal a significant 24.8% increase in overall weekly rides for users who had robotaxi rides. Our further analysis indicates a boost in demand for both robotaxi and regular ride-hailing services. This surge is particularly pronounced in robotaxi-accessible areas, whereas demand declines in inaccessible areas. Moreover, we find differential impacts on single-homing and multi-homing consumers, highlighting the nuanced implications for market competition. For single-homing consumers, robotaxi usage boosts demand in service-accessible areas, without significant changes elsewhere. Conversely, multi-homing consumers experience a more pronounced increase in demand in robotaxi-accessible areas (compared to single-homing consumers), whereas they significantly reduce their usage of the focal platform in inaccessible areas.

3 - Iv Estimation with Violated Exclusion Restrictions in Randomized Encouragement Designs

Cheng Lu, Washington University in St. Louis, St. Louis, MO, United States, Guangying Chen, Tat Chan, Zhengling Qi, Dennis Zhang

This paper examines the use of randomized encouragement designs as instrumental variables (IVs) for estimating treatment effects, despite potential violations of the exclusion restriction. Such violations occur when the encouragement design directly affects the outcome, beyond its indirect effect through the treatment. We demonstrate that, under conditions of heteroskedastic treatment intensity across groups, randomized encouragement designs can still yield unbiased estimations of treatment effects. Additionally, this approach allows for the isolation and measurement of the direct effect of the encouragement. Our simulation study illustrates that, in scenarios where the exclusion restriction is not met, our methodology recovers the true average treatment effects (ATEs) of both the endogenous treatment and the encouragement design, in contrast to the biased estimates produced by ordinary least squares (OLS) and two-stage least squares (2SLS). Furthermore, we introduce an adversarial min-max estimator to assess heterogeneous treatment effects (HTEs) of both the treatment and the encouragement design. We also propose a procedure for researchers to test the validity of the exclusion restriction in their encouragement designs by comparing the 2SLS estimates with those derived from our method. This procedure is demonstrated through an empirical study analyzing the impact of network size on the usage time of social media users, highlighting the practical importance of verifying the IV's exclusion restriction using our approach.

4 - Music Motion Synergy: a Multi-Modal Deep Learning Measure for Unveiling Engagement and Consumer Insights in Social Media Videos

Erya Ouyang, Temple University, Philadelphia, PA, United States, Wen Wang, Xueming Luo

Short-form videos are rapidly gaining popularity on social media as the leading format for entertainment and marketing. In this study, we propose a novel deep learning-based measure, referred to as Music Motion Synergy (MM Synergy), for assessing video quality on social media. We demonstrate that MM Synergy addresses three main challenges in video quality assessment: 1) leveraging core video content, 2) understanding interactions between multimodalities, and 3) integrating various engagement metrics that reflect management-relevant indicators. To derive this measure, we develop a new Multi-Task Multimodal Transformer model (MulT-MTL) and our extensive experimental results show that MulT-MTL achieves significantly lower forecast errors compared to various state-of-the-art deep learning models. More importantly, we show the practical value and superiority of MM Synergy as a general measure for assessing video quality and how MM Synergy can assist different influencers in creating better content. For example, we demonstrate that this measure not only informs about future multifaceted customer engagement but also predicts more nuanced psychometric consumer responses, providing valuable insights for stakeholders. This indicates that our measure can serve as quality check before publishing videos, or it can be embedded in recommendation systems or search engines to enhance the consumer experience. Furthermore, we investigate the mechanisms by quantifying the importance of capturing the key elements of MM Synergy. In summary, our methodology offers a decision-support measure to assist influencers, brands, and platforms in enhancing business outcomes and customer experiences of social media videos in this burgeoning entertainment market.

WB64

Regency - 602

Persuasion and Opinion Dynamics

Invited Session

Social Media Analytics

Chair: Tauhid Zaman, Yale University, New Haven, CT, United States

1 - Multi-Agent Nudging in Social Networks

Yen-Shao Chen, Yale University, New Haven, CT, United States, Tauhid Zaman

Influence campaigns on online social networks, run by organizations, political parties, and nation states, aim to sway large audiences using agents that disseminate persuasive content. However, their effectiveness is often limited by the bounded confidence phenomenon, where individuals are only influenced by viewpoints within a narrow spectrum. Our research demonstrates that agents must employ a strategy of gradual opinion shifts, or 'nudging', to successfully persuade under these conditions. We utilize control theory principles to develop a nudging policy and target selection strategy for multiple agents in an influence campaign. Simulations using real Twitter data reveal that these multi-agent nudging policies can effectively shift mean opinion and modify levels of opinion polarization. Our findings indicate that nudging strategies surpass other common techniques that overlook bounded confidence effects. Nonetheless, these strategies raise significant ethical concerns and risks of malicious use, emphasizing the need for careful consideration in the implementation of influence campaigns.

2 - Opinion Dynamics and Bayesian Persuasion

Tauhid Zaman, Yale University, New Haven, CT, United States

In this work, we demonstrate how various opinion dynamics models can be unified under a common Bayesian persuasion framework. We start with the assumption that an individual holds a prior belief about a certain state and receives signals regarding this state, with the likelihood of these signals being state-dependent. Within this Bayesian framework, these opinion dynamics models track the evolution of the posterior mean based on the received signals. By selecting appropriate priors and likelihoods, we are able to replicate many classic opinion dynamics models. We provide Bayesian formulations for the classic DeGroot model, the bounded confidence model, over-reaction, bounded shift, and the backfire effect. Additionally, we show how the structure of each opinion dynamics model is determined by the score (also known as the informant) of the signal likelihood.

3 - Investigating the Cumulative and Curvilinear Impact of Politicians' Negative Expressions on Twitter

Lingshu Hu, Washington and Lee University, Lexington, VA, United States, Kennon Sheldon, Rui Sun

The capacity to capture attention on social media holds paramount importance for public figures, as these platforms serve as pivotal arenas for exerting societal influence and accruing social and economic advantages. Particularly in political spheres, the ability to garner attention on social media not only mirrors politicians' social capital, likability, and popularity, but also shapes the communicative landscape of public forums, wielding influence over public attention and opinions, thus impacting our deliberative democracy. While the utilization of different sentiments to attract public attention has been extensively explored across social media and traditional media, previous studies have predominantly treated sentiment effects as linear phenomena, focusing on individual message units and overlooking the cumulative and non-linear dynamics of information dissemination. To address this gap, our study endeavors to employ computational and deep learning models to investigate the cumulative and curvilinear impacts of politicians' negative expressions using a comprehensive dataset sourced from Twitter. Moreover, individuals with diverse political identities often harbor contrasting worldviews and attitudes towards one another. Our research seeks to delve into how political partisanship and intergroup communication— encompassing dialogues with members of one's own political faction and those of opposing parties—moderate the influence of sentiment.

4 - The Highest Subreddit in the Land: Studying the Supreme Court Watchers Community on Reddit

Li Zhang, University of Virginia, Charlottesvle, VA, United States

This paper examines the vibrant community of Supreme Court watchers on Reddit, specifically within the "supremecourt" subreddit. By collecting a novel dataset of posts and comments, we analyze the discourse, engagement patterns, and topical trends among users. This study offers insights to understand how Reddit serves as a platform for public legal discussion, the nature of the conversations surrounding Supreme Court decisions and its Justices, and the characteristics of these participants.

WB65

Regency - 603

Causal Inference and Empirical Topics in (Digital) Markets

Invited Session

Information Systems

Chair: Dominik Molitor, Fordham University, New York, Germany

1 - The Impact of Uncertainty on Customer Satisfaction

Dominik Molitor, Fordham University, New York, Germany

This paper introduces a theoretical model of customer satisfaction that combines both a reference-dependent and a reference-independent component. Reference points are determined by expectations, which reflect the distribution of individual beliefs about performance outcomes. According to the model, uncertainty shifts subjective reference points, thereby increasing the level of performance at which customers start gaining satisfaction from reference dependence. The model further suggests that uncertainty mitigates the effect of both positive and negative deviations of actual performance from subjective reference points. Consequently, expected poor performance becomes less disappointing, while expected good performance becomes less gratifying compared to unexpected outcomes. The model's predictions are supported by findings from an experimental study on product delivery and a field study using data from an online retailer. Additionally, the paper features a web-based app that simulates the effect of uncertainty on customer satisfaction in a variety of customizable scenarios. The findings of the paper hold significant implications for firms' communication, customer valuation, and service strategies.

2 - Helping Small Businesses become more Data-Driven: A Field Experiment on eBay

Sagit Bar-Gill, Tel Aviv University, Tel Aviv, Israel, Erik Brynjolfsson, Nir Hak

As digitization sweeps across industries and markets, analytics and data-driven decision-making (DDD) are becoming increasingly important. The adoption of analytics and DDD has been slower in small-to-medium enterprises (SMEs) compared to large firms, and reliable causal estimates of the impacts of analytics tools for small businesses have been lacking. We derive experiment-based estimates of the impact of an analytics tool on SME outcomes, analyzing the randomized introduction of eBay's Seller Hub (SH), a data-rich seller dashboard. We find that SH adoption is associated with increased DDD, and that access to SH increases e-retailers' revenues by 3.6% on average, as more items are transacted and service quality increases, without increases in average prices. Our results suggest that analytics and DDD help SMEs establish a competitive advantage. Managerial practices play an important role in reaping the benefits from the analytics dashboard, as over a third of the SH impact is driven by active performance monitoring. Digital platforms should, therefore, embed analytics tools to support SMEs, to increase entry and revenues. Furthermore, policies to support small businesses' transition to the data era should address analytics and DDD gaps by ensuring access to both tools like SH and to appropriate managerial training.

3 - Analyzing Store Cannibalization Using Mobility Data

Shachar Reichman, Tel Aviv University, Coller School of Management, Tel Aviv, Israel, Peter Zubcsek, Tal Shoshani

Retailers commonly use mobile location data to improve marketing campaigns, while other uses of consumer location data in retail analytics are limited. For instance, although the literature has recognized the importance of mapping store cannibalization of brick-and-mortar outlets, prior work in this area has primarily attempted to quantify it based on consumers' distance from stores. This research presents a mobile location analytics application to model store cannibalization, using the specialty coffee retail market as a testbed. Specifically, we addressed: (1) How can mobile location data capture the externalities of store cannibalization? (2) What is the predictive power of urban mobility data regarding store performance? (3) How does the predictive performance of semi-supervised models using aggregate mobility data compare to supervised models with network features from individual mobility data? We answered these questions using two large sets of urban mobility data, focusing on co-visitations (when a consumer visits two stores within a week). We constructed a network of retail stores, linking nearby outlets with weights proportional to co-visitations and the inverse of the squared distance between stores. Using network analysis techniques and a graph transformer model, we extracted store-level features to predict store closures. Preliminary results indicate that analyzing store co-visitation networks improves model performance in assessing store health compared to using only census-based economic variables. Our research contributes to management and marketing literature by showing how consumer mobility data can help understand and predict store cannibalization dynamics.

4 - User Data and Brokers

Joy Wu, UBC Sauder School of Business, Vancouver, BC, Canada, Ananya Sen

Data brokers sell access to and usage of personal information they harvest about individuals and facilitate the exchange of personal data utilized throughout the digital economy. However, little is known about people's valuation, perceptions, and demand for their privacy from data brokers. We conducted a survey experiment of 4,000 U.S.-representative individuals and provisioned information about how brokers harvest user data from two broad categories of data transfers people participate in: those with government agencies and those with commercial entities. We gathered individuals' beliefs about data exposure and their revealed preferences for privacy from brokers. Perceptions about exposure to brokers shift in a direction that suggests Americans largely underestimate their data exposure—particularly the size of the broker market that harvests government records—and these beliefs are malleable to information interventions. Despite shifting their beliefs, individuals' willingness to pay to delete data from brokers is unaffected by our interventions on the aggregate level. However, we find significant variations in how effective information interventions are at influencing valuations for privacy based on characteristics of individuals, such as whether they conducted prior data deletion requests and whether they have applied for a change of address request with USPS. These results highlight the importance of how, even though individuals can be systematically underinformed about how much privacy they have, their privacy choices depend on their idiosyncratic, instrumental value of privacy.

WB66

Regency - 604

Enhancement of Object Detection

Invited Session

Artificial Intelligence

Chair: Nicolas Bustos, University of South Florida, Plant City, FL, 33613, United States

1 - FT-Cyclegan: A Novel Frequency-Based Architecture for Visual-and-IR Image Translation to Enhance Multimodal Object Detection

Nicolas Bustos, University of South Florida, TAMPA, FL, United States

Multimodal architectures are being touted as the future of object detection. These architectures can overcome environmental challenges such as adverse lighting and weather conditions to deliver more accurate detection. The fusion of visual and infrared (IR) data has been shown in the literature to improve the detection under fog and poor lighting conditions. However, there is a lack of publicly available multimodal datasets that are paired and aligned for use in multimodal models and data fusion algorithms. In this work, we introduce FT-CycleGAN, a CycleGAN-based architecture that uses Fourier transform (FT) within the computation of its cycle consistency loss. We generate synthetic images using FT-CycleGAN that multimodal object detection algorithms can leverage to improve their performance. The FT component of the cycle consistency loss is computed by matching the frequency amplitude between the input and cycled images. FT-CycleGAN is used to generate visible-and-infrared translated images from the Teledyne FLIR dataset. Faster-RCNN is shown to perform better utilizing the generated image pairs.

2 - The Impact of using Data Fusion with Synthetic Images on Multimodal Object Detection

Mehrsa Mashhadi, University of South Florida, Tampa, FL, United States, Nicolas Bustos, Susana Lai-Yuen, Tapas Das

The performance of object detection models can be compromised under adverse weather conditions such as reduced visibility and altered lighting. Models trained on unimodal data struggle, as these conditions distort visual data. Specifically, models relying solely on visible imagery, while rich in texture details, are vulnerable to weather variations, compromising their effectiveness. Thermal images offer a robust alternative since they are less affected by lighting. Combining visible and thermal imagery enhances detection robustness by leveraging both modalities. Despite various fusion approaches, comprehensive evaluations of visible and thermal imagery for object detection on diverse weather conditions are limited due to a lack of standardized benchmarks and the publicly available datasets containing paired images with weather labels. We used data fusion with synthetic images utilizing the Teledyne FLIR dataset, which features unpaired visible and thermal images labeled with weather conditions. Synthetic pairs were generated using CycleGAN. Advanced fusion techniques, SeAFusion and GAN-FM, were applied, and models such as Faster R-CNN, YOLOv7, and Detection Transformers were assessed across different weather scenarios and object sizes.

Results indicate that training Faster R-CNN with thermal images provides the best results across different scenarios. The use of synthetic data with fusion techniques generally yields lower results compared to using unimodal images only. Comparing the data fusion techniques, our findings suggest that using fused images with SeAFusion improves detection performance compared to GAN-FM. These insights highlight the potential benefits of incorporating synthetic and multimodal data and offer valuable considerations for improving object detection models under diverse weather conditions.

3 - Zero-Shot Vision-Language Models: Advancing Multimodal Understanding

Raphael Caleb Yao, University of San Francisco, San Francisco, CA, United States

The emergence of zero-shot vision-language models has marked a pivotal development in the field of artificial intelligence, enabling models to comprehend and interpret visual data with minimal to no direct task-specific training. These models combine the strengths of natural language processing and computer vision, allowing for the understanding of images and text in tandem. CLIP (Contrastive Language-Image Pretraining) utilizes large-scale training on diverse image-text pairs to learn generalized representations. This approach allows it to recognize visual concepts based solely on textual descriptions, demonstrating remarkable zero-shot transfer capabilities. Similar models like DALL-E and ALIGN have shown that the integration of language and vision enables complex tasks such as image generation from text prompts. This paper explores architecture and training strategies behind a zero-shot vision-language model that it develops, examining its potential for applications like image classification, object detection, and multimodal search without task-specific fine-tuning. It analyzes the challenges involved in understanding visual concepts from textual descriptions, the models' generalization abilities across domains, and the implications for future research in multi-modal learning.

WB67

Regency - 605

AI and Optimization in Manufacturing

Contributed Session

Chair: Zhaoguang Xu, Dalian University of Technology, Dalian City, 116024

1 - Evaluating Machine Learning and Multi-variate Empirical Models in Solder Joint Characteristic Life Prediction

Seyed Soroosh Alavi, Auburn University, Auburn, AL, United States, Dr. Daniel F. Silva

This research compares the application of machine learning (ML) vs. a newly developed multi-variate physics-based empirical model to predict the characteristic life of solder joints, previously estimated through single-variable empirical models. Both approaches utilize data from isothermal aging tests followed by thermal cycling tests conducted on solder joints. The ML approach seeks to create a universal prediction model that consolidates various test conditions. Concurrently, the empirical model incorporates and uses both aging duration and temperature as input variables to provide a robust alternative to traditional single-variable empirical models, which require separate models for each aging condition. Comparative analysis shows that both the ML and the two-variable empirical models achieve high and comparable accuracy to traditional single-variable empirical models. The ML approach offers significant cost and time efficiencies, while the two-variable empirical model demonstrates an ability to combine test conditions, reducing the number of necessary models and experiments. The final comparative study between both models discusses the advantages and disadvantages of each, highlighting their practical applications and performance merits in the field of solder joint characteristic life prediction.

2 - Defect Detection for OHT Wheels in Semiconductor Manufacturing Processes Using Generative AI-based Data Augmentation

Chang Hyun Lee, Department of Industrial Engineering, Sungkyunkwan University, Suwon, Korea, Republic of, JUHYUN KIM, Sungjun Hur, Dong Hee Lee

Monitoring the wheel condition of Overhead Hoist Transport (OHT) systems is crucial for maintaining production efficiency and safety in semiconductor manufacturing processes. However, collecting defect data in real manufacturing environments is challenging, and the defects often appear in very small regions within the images, limiting the performance of existing machine learning models. This study proposes a method to train robust defect detection models from limited real data by leveraging generative AI technology to augment OHT wheel defect images.

The proposed technique first generates various virtual defect patterns resembling real defects and seamlessly synthesizes them onto OHT wheel images using appropriate image processing techniques, creating realistic defective wheel images. Domain adaptation techniques are then applied to minimize the domain gap between the generated synthetic images and real images. The augmented dataset is utilized to train a convolutional neural network-based defect detection model, and its performance is evaluated in an actual manufacturing environment.

Experimental results demonstrate that the proposed method achieves higher accuracy compared to models trained solely on real data, highlighting the effectiveness of generative AI-based data augmentation in addressing the limited real data problem.

3 - A Heuristic Approach to the Problem of Min-Time Coverage in Constricted Environments with Arbitrary Guidepath Networks

Young In Kim, ISyE Georgia Tech, Atlanta, GA, United States, Spiridon Reveliotis

In a recent work, we investigated some min-time coverage problems that use networked robotic fleets for conducting routine inspections and other monitoring services in well-structured but constricted environments. That work provided a detailed description of these problems, their formal characterization in the form of a Mixed Integer Programming (MIP) formulation, and strong combinatorial relaxations that can significantly reduce the number of integer variables without compromising the correctness of the derived solutions with respect to the original formulations. The current work develops a heuristic approach to address larger problem instances that are not amenable to the aforementioned relaxations. Numerical studies demonstrate the efficacy and the computational gains achieved by the proposed heuristic.

4 - Optimization-Informed Supervised and Unsupervised Learning for Energy-Cost and Thermal Comfort-Aware Air Conditioning

Minseo Kim, Yonsei University, Seoul, Korea, Republic of, Soongeol Kwon

This study aims to develop an optimization method for scheduling AC operations to minimize energy costs while ensuring indoor thermal comfort. We propose an optimization-informed learning method that integrates unsupervised and supervised learning models with a mathematical optimization model. This integration allows the mathematical optimization model to interpret data from an optimization perspective, enhancing the understanding of uncertainties and variabilities. The method begins by clustering uncertain weather and time-variable price data to identify underlying patterns. A two-stage chance-constrained stochastic program then defines AC operation schedules for each cluster. Finally, the most appropriate schedule is selected through supervised learning, considering uncertainties and variabilities in the clustered data. This process enables the optimization model to provide the learning models with an informative form of data and criteria to ultimately determine proper AC operations in an energy-cost and thermal comfort-aware scheme.

5 - Optimization of Multi-Period Quality Costs Leveraging Predictive Analytics in Intelligent Manufacturing

Zhaoguang Xu, Dalian University of Technology, Dalian, China, People's Republic of, Stefan Minner

The trade-off between quality and cost is a key challenge faced by manufacturing enterprises. In the environment of intelligent manufacturing, predictive quality technology offers new opportunities for improving quality and reducing costs. However, it is difficult to use predictive results to optimize quality costs quantitatively. Thus, we propose a predict-then-optimize approach for quality cost optimization. In particular, we first leverage industrial big data and machine learning techniques to construct a deep learning-based quality prediction model to estimate the defect probabilities in products. These predictive insights are then incorporated into a multi-period knapsack problem model for the second stage to optimize the total quality costs, considering the tradeoff between inspection and external failure costs. Considering the false negative and false positive errors of the prediction models, we analyze the mechanism of impact of the predictive model on inspection decisions and quality costs.

WB68

Regency - 606

AI in Action: Enhancing Decision-Making and Collaboration

Contributed Session

Chair: Michelle Vaccaro, MIT, Cambridge, MA, United States

1 - When Are Combinations of Humans and AI Useful?

Michelle Vaccaro, MIT, Cambridge, MA, United States, Abdullah Almaatouq, Thomas Malone

Inspired by the increasing use of AI to augment humans, researchers have studied human-AI systems involving different tasks, systems, and populations. Despite such a large body of work, we lack a broad conceptual understanding of when combinations of humans and AI are better than either alone. Here, we addressed this question by conducting a meta-analysis of over 100 recent experimental studies reporting over 300 effect sizes. First, we found that, on average, human-AI combinations performed significantly worse than the best of humans or AI alone. Second, we found performance losses in tasks that involved making decisions and significantly greater gains in tasks that involved creating content. Finally, when humans outperformed AI alone, we found performance gains in the combination, but when the AI outperformed humans alone we found losses. These findings highlight the heterogeneity of the effects of human-AI collaboration and point to promising avenues for improving human-AI systems.

2 - Evaluate AI or Evaluate Myself? Effects of AI Explanations on Appropriate Decision Adjustment

Mandie Liu, City University of Hong Kong, Hong Kong, Hong Kong, Weiling Ke, David Xu

To enhance interpretability and mitigate human resistance, AI-advised decision-making increasingly comes with explanations about why they make the predictions. However, the effects of AI explanations on appropriate decision adjustment have yielded mixed results. Therefore, it’s urgent to comprehensively understand how AI explanations influence human decision-making process. Drawing on dual process theory, our research addressed this question by examining the role of AI explanations in deliberate processing, specifically self-reflection, which is a pivotal aspect of effective decision-making. This approach differs from prior studies, which focused on heuristic processing, like perceived expertise in AI. Employing the context-specific framework, we also theorize how AI explanation, as a technology-related variable, interacts with other context aspects, including a user characteristic (cognitive bias) and usage context variable (advice distance). Through two online experiments (N=120 and N=366) conducted in the context of stock investment, we found that 1) AI explanations bolster both perceived expertise and self-reflection, but only when advice distance is far; 2) compared to perceived expertise, self-reflection had stronger effects on hit rates but weaker effects on false alarms, which indicates that evaluation for initial judgment, rather than evaluation for AI, more directs appropriate decision adjustment. 3) when users suffer from cognitive bias (specifically familiarity and herding biases in our study), AI explanations showed stronger effects on driving self-reflection when advice distance is far. Our study advances the understanding of the mechanisms underlying the relationship between AI explanations and appropriate decision adjustment. We also offer practical insights for AI providers to better support effective decision-making.

3 - An Empirical Study on the Methods for an Effective Collective-Interest AI Recommendation Systems

Yalda Hojjati, Auburn University, Auburn, AL, United States, Yuanyuan Chen, Uzma Raja

AI recommendation systems have become ubiquitous, shaping various facets of modern life. While much research has focused on improving the effectiveness of individual-centric AI recommendation systems, fewer studies have examined AI systems providing recommendations based on collective interests. Collective-interest recommendation systems hold promise in a wide range of domains, such as public health emergency management, climate change, transportation optimization, and community resource allocation. Unlike their individual-centric counterparts, collective-interest recommendation systems offer suggestions to benefit groups rather than serve personal interests. Advancing collective interests requires individuals to engage in cooperative and altruistic behaviors, often at a personal cost for the greater good. However, the success of these systems depends heavily on user adoption. Achieving collective interest will become a formidable challenge if users do not trust or adopt AI recommendations. Therefore, investigating methods to promote user trust and adoption of AI-recommended cooperative and altruistic behavior aiming for collective interest is theoretically and managerially important.

Our research examines the mechanisms for designing AI recommendation systems that foster trust in AI technology and encourage individuals to adopt collective interest recommendations. Specifically, we aim to evaluate the effectiveness of different modalities and explore the impact of elucidating the AI's underlying process on user trust and adoption of such recommendations. The study also has significant managerial implications, as it helps organizations design systems that promote collective interests while fostering user trust and engagement.

4 - How Forced Intervention Facilitates Long-term Algorithm Adoption

Jiankun Sun, Imperial College London, London, United Kingdom, Xinyu Cao, Chenshan Hu, Dennis Zhang

While artificial intelligence (AI) technologies increasingly become powerful and useful in daily operations, human workers often resist adopting algorithmic recommendations, a phenomenon known as algorithm aversion. This aversion can undermine the performance of algorithms in practical applications. In this paper, we investigate whether and why forced interventions can promote algorithm adoption and reduce algorithm aversion in the long term. Data from a leading online education company reveal that sales workers underutilize a new matching algorithm and often apply it to low-quality leads. A field experiment was conducted where sales workers were required to use or abstain from using the algorithm for three weeks. Experimental results show that forcing workers to use the algorithm during the experiment causally increases their algorithm usage one month after the experiment by 15.8 percentage points. We develop a theoretical model to explore the mechanisms behind this improvement. Our findings suggest that learning is the primary mechanism driving algorithm adoption among workers. Enforced usage allows workers to experience the algorithm's unbiased performance firsthand and positively adjust their beliefs about it. Consequently, after the experiment, the workers in the treatment group not only use the algorithm more frequently but are also more likely to apply it to high-quality leads. The study offers empirical evidence that forced intervention can effectively enhance long-term algorithm adoption among workers. It also suggests that firms could implement extrinsic interventions or educational programs to help workers recognize the benefits of algorithms and develop unbiased beliefs about their capabilities, thus promoting sustained algorithm usage.

5 - AI Testing as a Critical Milestone towards Effective AI Integration in Serious Games Test-Beds for Human-AI Collaboration

Amir Abdollahi, Northeastern University, Boston, MA, United States

As AI becomes more prevalent in our lives, effectively integrating it and determining the best conditions for human-AI collaboration is crucial, especially in critical areas like public policy and healthcare. We introduce a framework for complex policymaking, focusing on pandemic management to simulate real-world scenarios, deepening our understanding of AI integration in this essential context. By gamifying the policymaking process, we simulate real-world scenarios, aiming to validate AI's effectiveness before actual deployment in human-AI collaboration contexts.

Utilizing deep reinforcement learning (DRL), we train an agent on historical lockdown and travel restriction data from 186 countries during the COVID-19 pandemic. Comparing the DRL agent's performance with governmental actions, we evaluate AI-driven policies' potential impact on infection rates, fatalities, and recoveries. Results show the DRL agent's adeptness in understanding epidemic dynamics, often surpassing governments' performance by reducing fatalities and infections while increasing recovery rates. However, our initial study encountered limitations as it relied on historical data, hindering the construction of an entirely realistic counterfactual pandemic scenario.

Consequently, we created a more comprehensive simulated environment for better generalization. Our approach, incorporating the SEIRD model, demonstrates the DRL agent's effectiveness in pandemic management, improving recovery rates and reducing infection and mortality rates over time. This research emphasizes the importance of assessing AI performance before implementing human-AI collaboration in sensitive contexts. We propose future research to expand on this work by incorporating human-AI collaboration scenarios and considering financial constraints, providing a holistic perspective on efficient pandemic policymaking.

6 - Travel Demand Forecasting: A Fair AI Approach

Xiaojian Zhang, University of Florida, Gainesville, FL, United States, Qian Ke, Xilei Zhao

Artificial Intelligence (AI) and machine learning have been increasingly adopted for travel demand forecasting. While AI-based models offer precise predictions, they can also introduce biases and fairness concerns. Relying on these biased models for decision-making could worsen social inequalities in transportation policies. Therefore, in this study, we propose a novel methodology to develop fairness-aware, highly-accurate travel demand forecasting models. Particularly, the proposed methodology can enhance the fairness of AI models for multiple protected attributes (such as race and income) simultaneously. Specifically, we introduce a new fairness regularization term, which is explicitly designed to measure the correlation between prediction accuracy and multiple protected attributes, into the loss function of the travel demand forecasting model. We conduct two case studies to evaluate the performance of the proposed methodology using real-world ridesourcing-trip data in Chicago, IL and Austin, TX, respectively. Results highlight that our proposed methodology can effectively enhance fairness for multiple protected attributes while preserving prediction accuracy. Additionally, we have compared our methodology with three state-of-the-art methods that adopt the regularization term approach, and the results demonstrate that our approach significantly outperforms them in both preserving prediction accuracy and enhancing fairness. This study can provide transportation professionals with a new tool to achieve fair and accurate travel demand forecasting.

WB69

Regency - 607

Hotelling Medal Lecture

Invited Session

Energy: Natural Resources

Chair: Ben Hobbs, Johns Hopkins University, Baltimore, MD, United States

Co-Chair: Shmuel Oren, UC Berkeley, Berkeley, United States

1 - 2024 Hotelling Medal Recepient

2 - Mathematically Based Global Optimization Algorithms and their Application to Computationally Expensive Environmental Pollution Problems

Christine Shoemaker, Cornell University, Singapore, Singapore

There are many environmental problems that could benefit from optimization analysis, but their solution is challenging because the problems are not convex and the evaluation of the objective function (usually an expensive multimodal numerical simulation model) allows only a modest number of objective function evaluations. Shoemaker's group have developed a series of mathematically based global optimization algorithms with radial basis function surrogates that can be efficiently used with costly simulation models to identify cost-effective pollution control and climate change predictions . These “PYSOT” algorithms are freely available online. Application to very expensive PDE models of lake pollution and of global climate change will be presented.

WB70

Regency - 701

Harnessing Machine Learning and Advanced Optimization for Power System Planning in Developing Countries

Invited Session

ENRE: Electricity

Chair: Yifu Ding, MIT Energy Initiative, Cambridge, MA, United States

1 - Optimizing the least-cost decarbonization pathways for India's power system through data-driven coal plant retrofitting strategies

Yifu Ding, MIT Energy Initiative, Cambridge, MA, United States, Dharik Mallapragada, Serena Patel, Jansen Wong, Guiyan Zang, Robert Stoner

India set two Nationally Determined Contribution (NDC) targets by 2030 to achieve the net-zero carbon emission goal by 2070, which requires deep decarbonization of India’s power generation sector. Yet in India, coal-fired electricity contributes to more than 60% of power generation, and the government still permits new construction and life extension of coal plants. This study creates a dataset for the operating characteristics of 806 India's coal-fire power units using machine learning. Then we optimize the least-cost decarbonization pathways for India's power system considering two retrofitting technologies: carbon capture and storage (CCS) and biomass co-firing. Results show the optimal coal retrofitting strategy could lead to a more balanced distribution of generation capacity, decrease the unabated coal plant capacity, and lower the electricity generation and carbon abatement costs under a relatively stringent carbon cap of less than 1000 Mt CO_2 to meet India's NDC goals.

2 - Location-based regional power decarbonization planning for a just energy transition: exploring coal repurposing opportunities

Ziting Huang, Johns Hopkins University, Baltimore, MD, United States, Benjamin Hobbs, Debabrata Chattopadhyay, Qingyu Xu

Deep decarbonization of energy system may give rise to stranded asset risks for some regions where thermal power dominates the energy mix. Regarding coal power assets, we put forward five alternatives to assist with energy decarbonization: demolishing completely, mothballing for backup generation, retrofitting with CCS, repurposing with renewable energy for cleaner generation, and redeveloping into other economic activity use. Each alternative has different levels of asset reutilization, thus varying costs and benefits in generation, local economy, and environment. Such trade-offs between macro decarbonization goals and local socio-economic changes are not well captured in the energy system planning.

This study aims to co-optimize goals in reducing GHG emissions while achieving societal benefits in local communities by exploring the future role of coal in a regional energy system. Through an agent-based model, which simulates the interaction between coal agent and power system environment, the study identifies the optimal decarbonization strategy for existing coal assets at both plant and system level. With the locational information on socioeconomics, this analysis framework incorporates societal impacts from coal power’s decisions in the regional electricity planning, which provides valuable implications for a just energy transition. As a comparison, the study further explores the performance differences between an agent-based model and a generalized multi-objective optimization. The methodology is applied in China and South Africa for case study illustration, where coal is a major electricity source.

3 - Supply Security Under Institutional Constraints and the Low-Carbon Energy Transition in China

Ming Wei, University of California San Diego, San Diego, CA, United States, Michael Davidson

In the context of meeting China's climate change targets, a massive scale up of low-carbon energy is required while meeting growing demand, presenting critical challenges for maintaining electric supply security. Institutionally, China’s power sector—as with many other systems—is also in a period of transition from central planning toward market liberalization. This study presents a retrospective analysis of the power shortages that plagued China from 2020 to 2022, examining their causes, commonalities, impacts, and governmental responses. Modeling simulations using unit commitment and economic dispatch models are then conducted to investigate the underlying operational and institutional factors. These analyses rely on extensive collection of publicly available datasets, ensuring transparency and reproducibility. The results highlight the inflexibilities of existing market mechanisms and institutional arrangements in adapting to the demands of a decarbonizing energy system. The study also explores forward-looking scenarios through 21 simulations incorporating extreme weather conditions, electricity and thermal coal pricing dynamics, and different dispatch mechanisms. These scenarios reveal critical market and institutional vulnerabilities, suggesting that reforms to improve supply security must consider a system-wide perspective beyond standard approaches such as increasing capacity. The results contribute to the ongoing discourse on effective strategies for advancing China's energy transition.

4 - Planar location-allocation problems for spatial planning of decentralized energy systems

Beste Akbaş, Imperial College London, London, United Kingdom, Ayse Selin Kocaman

One major impediment to developing countries’ economic growth is the lack of access to affordable, sustainable, and reliable modern energy systems. Hundreds of millions of people live in rural areas and do not have access to essential electricity services. In this study, we present a planar facility location–allocation problem for planning decentralized energy systems in rural development. We consider nano-grid and micro-grid systems to electrify rural households. While micro-grids serve multiple households with a common generation facility, nano-grids are small-scale systems serving individual consumers. The households served by micro-grids are connected to the generation facilities with low-voltage cables, for which we employ a distance limit constraint due to technical concerns, including power loss and allowable voltage levels. In this problem, we minimize the total investment cost that consists of the facility opening and the low-voltage cable costs. In order to capture the diversity of cost structures in renewable energy investments, we consider three versions of the objective function where we incorporate different combinations of fixed and variable cost components for facilities. For this problem, we provide mixed-integer quadratically constrained problem formulations and propose model-based and clustering-based heuristic approaches. Model-based approaches are multi-stage, in which we solve the discrete counterparts of the problem and employ alternative selection methods for the candidate facility locations. Clustering-based approaches utilize faster clustering techniques to identify the type and location of the facilities. We conduct computational experiments on real-life instances from villages in Sub-Saharan Africa and perform a comparative analysis of the suggested heuristic approaches.

WB71

Regency - 702

Navigating Challenges in Statistical Inference and Decision-Making: Adaptive Experiments Design, Testing Complex Hypotheses, Policy Optimization, and Robust Mendelian Randomization

Invited Session

Data Mining

Chair: Mengxin Yu, University of Pennsylvania, Philadelphia, PA, 19104, United States

1 - An Online Optimization Approach to Designing Adaptive Experiments

Christopher Harshaw, MIT / UC Berkeley, Cambridge, MA, United States

From clinical trials to development economics, randomized experiments are increasingly adopted to estimate causal effects. Adaptive experiment designs – where experimental subjects arrive sequentially and the probability of treatment assignment can depend on previously observed outcomes – are becoming an increasingly popular method for causal inference, as they offer the possibility of improved precision over their non-adaptive counterparts. However, in simple settings (e.g. two treatments) the extent to which adaptive designs can improve precision is not sufficiently well understood. In this talk, I present my recent work on the problem of Adaptive Neyman Allocation, where the experimenter seeks to construct an adaptive design which is nearly as efficient as the optimal (but infeasible) non-adaptive Neyman design which has access to all potential outcomes. I will show that the experimental design problem is equivalent to an adversarial online convex optimization problem, suggesting that any solution must exhibit some amount of algorithmic sophistication. Next, I present Clip-OGD, an experimental design that combines the online gradient descent principle with a new time-varying probability-clipping technique. I will show that the Neyman variance is attained in large samples by showing that the expected regret of the online optimization problem is bounded by O(\sqrt{T}), up to sub-polynomial factors. Even though the design is adaptive, we construct a consistent (conservative) estimator for the variance, which facilitates the development of valid confidence intervals. I will conclude with recent progress on extending this work to covariate-adjusted estimators and covariate-responsive designs, which is made possible through the online optimization perspective.

2 - Hunt, Test and Aggregate: a Flexible Framework for Testing Complex Hypotheses

Richard Guo, University of Washington, Seattle, WA, United States

Hypotheses arising from real-world problems can be rather complex: (1) the null hypothesis can be large so it is difficult to test directly; (2) the alternative can be large and heterogenous, so it is challenging to devise a test that has power against a variety of alternatives. However, if such a hypothesis can be represented as a conjunction of simpler hypotheses corresponding to different directions, "hunt and test" provides a useful strategy: split the data into A and B; use A to hunt for a direction that contains signal against the null; and use B to test the simpler hypothesis associated with that direction. Nevertheless, applying "hunt and test" alone can suffer from non-replicability and low power. To remedy these problems, we demonstrate that by properly aggregating and calibrating a large number of hunt-and-tests, we can obtain an ensemble procedure that is reliable, replicable, and highly powerful. We demonstrate how to use this framework to construct tests for a variety of problems, including detecting subtypes of cancer and testing specification of regression models.

3 - Policy Learning "Without" Overlap: Pessimism and Generalized Empirical Bernstein's Inequality

Ying Jin, Data Science Institute, Harvard University, Cambridge, MA, United States, Zhimei Ren, Zhuoran Yang, Zhaoran Wang

Offline policy learning aims at utilizing observations collected a priori (from either fixed or adaptively evolving behavior policies) to learn an optimal individualized decision rule that achieves the best overall outcomes for a given population. Existing policy learning methods rely on a uniform overlap assumption, i.e., the propensities of exploring all actions for all individual characteristics must be lower bounded. As one has no control over the data collection process, this assumption can be unrealistic in many situations, especially when the behavior policies are allowed to evolve over time with diminishing propensities for certain actions. In this work, we propose Pessimistic Policy Learning (PPL), a new algorithm that optimizes lower confidence bounds (LCBs) -- instead of point estimates -- of the policy values. Without assuming any uniform overlap condition, we establish a data-dependent upper bound for the suboptimality of our algorithm, which only depends on (i) the overlap for the optimal policy, and (ii) the complexity of the policy class we optimize over. As an implication, for adaptively collected data, we ensure efficient policy learning as long as the propensities for optimal actions are lower bounded over time, while those for suboptimal ones are allowed to diminish arbitrarily fast. In our theoretical analysis, we develop a new self-normalized type concentration inequality for inverse-propensity-weighting estimators, generalizing the well-known empirical Bernstein's inequality to unbounded and non-i.i.d. data. We complement our theory with an efficient optimization algorithm via Majorization-Minimization and policy tree search, as well as extensive experiments that demonstrate the efficacy of PPL.

4 - Genius-Mawii: for Robust Mendelian Randomization with Many Weak Invalid Instruments

Ting Ye, University of Washignton, Seattle, WA, United States

Mendelian randomization (MR) addresses causal questions by using genetic variants as instrumental variables. We propose a new MR method, GENIUS-MAWII, which simultaneously addresses the two salient challenges in MR: many weak instruments and widespread horizontal pleiotropy. Similar to MR-GENIUS, we use heteroscedasticity of the exposure to identify the treatment effect. We derive influence functions of the treatment effect, and then we construct a continuous updating estimator and establish its asymptotic properties under many weak invalid instruments asymptotic regime by developing novel semiparametric theory. We also provide a measure of weak identification, an overidentification test, and a graphical diagnostic tool.

WB72

Regency - 703

Bayesian Optimization

Invited Session

Data Mining

Chair: Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

Co-Chair: Peter Frazier, Cornell / Uber, Ithaca, NY, 14850, United States

Co-Chair: Chong Liu, State University of New York at Albany, Albany, NY, 12222, United States

1 - Bayesian Optimization for Scientific Discovery: the Value of Function Networks and Partial Evaluations

Poompol Buathong, Cornell University, Ithaca, NY, United States, Jiayue Wan, Samuel Daulton, Raul Astudillo, Maximilian Balandat, Peter Frazier

Scientific discovery, including the search for new materials, drugs, and technologies, presents exciting growth opportunities for operations research and management science methods. Bayesian optimization of function networks (BOFN) has emerged as a new grey-box BO framework, offering an attractive paradigm for guiding experiments in scientific discovery. BOFN models the objective of an application as a function network and leverages the intermediate information obtained from evaluating each node in the network to improve the sampling efficiency. Similar to standard BO, BOFN can handle expensive, noisy and non-convex objective functions and constraints where derivatives are not typically available. In this talk, we introduce a significant advancement in this area – guiding so-called partial evaluations in which only a part of the function network can be evaluated. This is critical for state of the art performance in many scientific discovery applications. We demonstrate the uses of the framework in two applications, namely, designing new catalysts for hydrogen fuel cells and optimizing experimental conditions for protein encapsulation. Our numerical results show that BOFN with partial evaluations outperformed existing methods, providing a better solution while spending less evaluation budget.

2 - Parametric Function Approximation for Black-Box Optimization: A New Perspective

Chong Liu, State University of New York at Albany, Albany, NY, United States, CHUANHAO LI, Yu-Xiang Wang

To discover strong materials, materials scientists need to sequentially select processing parameters and conduct long-time experiments to observe performances. To train a deep neural network, computer scientists need to run computationally expensive hyperparameter tuning experiments on validation sets. In both cases, scientists are solving black-box optimization problems. Existing work like Bayesian optimization usually assumes objective function is drawn from some Gaussian process (GP) which unfortunately usually suffers from the curse of dimensionality. In this talk, we provide a new perspective of solving such problems without GP but instead with parametric function approximation. Common acquisition functions like upper confidence bound (UCB) can still be constructed in parameter space to help us select next data points. Both regret bounds and empirical results will be shown. In the end, I'll show our recent progress of extension to federated optimization settings.

3 - Bayesian Deep Kernel Learning for Bayesian Optimization

Christopher Yeh, California Institute of Technology, Pasadena, CA, United States, James Bowden, Raul Astudillo, Yisong Yue

Bayesian optimization is a framework for optimizing functions with time-consuming evaluations. It has been successful in a broad range of applications, from hyperparameter tuning to chemical design. Bayesian optimization methods use a probabilistic surrogate model of the objective function, most commonly Gaussian processes. However, using a Gaussian process requires specifying a kernel. The choice of the kernel is critical for achieving good performance, but this choice is often challenging for practitioners. In this work, we propose using Bayesian deep kernel learning to automatically learn a kernel "on the fly". Intuitively, this approach learns a distribution over embeddings over which a simpler kernel (such as a linear kernel) can be used effectively. We show that our method outperforms standard approaches using RBF and Matern kernels as well as a non-Bayesian deep kernel learning approach in several synthetic and real-world test problems, including protein and nanophotonics design problems. We also show that Thompson sampling under this probabilistic model enjoys a sublinear regret guarantee.

4 - Bayesian Optimization of Conditional Gaussian Process Trees

Mengrui(Mina) Jiang, Arizona State University, Tempe, AZ, United States, Giulia Pedrielli

Bayesian optimization algorithms are used in many complex problems to solve expensive black-box functions without using any information about the structure or properties of the function. In some cases additional information about the system might be available, for example, the trajectory of a dynamic system, the intermediate execution of discrete states, etc. While Bayesian optimization has been successfully applied to many problems, there is no standard way to incorporate known structure. In this work, we specifically consider extending Bayesian optimization to loss functions whose heterogeneity structure can be encoded by a tree. In particular, the response function is a finite set of functions, and constraints at the edges are imposed on intermediate output associated with the non-leaf nodes to define which path the sample is assigned to. Such response function can be modeled through a conditional Gaussian process tree (CGPT). In this work, we propose for the first time a Bayesian optimization algorithm that designs a novel acquisition function in addition to the CGPT model as the surrogate, and investigate its performance in synthetic and industrial experiments. We call the resulting algorithm the Bayesian Optimization with Conditional Gaussian Process Trees (BOCGPT).

WB73

Regency - 704

Scaling ML applications with Federated Learning

Invited Session

Data Mining

Chair: Brianna Mueller, University of Iowa, Iowa City

Co-Chair: Zhenyu Sun, Northwestern University, Evanston, 60201, United States

1 - Dynamic Ensemble Selection for Personalized Federated Learning

Brianna Mueller, University of Iowa, Iowa City, IA, United States

Federated learning (FL) enables multiple clients with distributed data sources to collaboratively train a shared model without compromising data privacy. However, existing FL paradigms face challenges due to heterogeneity in client data distributions and system capabilities. Personalized federated learning (pFL) has been proposed to mitigate these problems, but often requires a shared model architecture and a central entity for parameter aggregation, resulting in scalability and communication issues. In this context, we introduce a novel approach that leverages dynamic ensemble selection (DES), which not only offers personalized models for individual clients but also delivers unique solutions for each test sample. Our proposed framework is based on peer-to-peer model sharing, where clients maintain a collection of models from all participants in the network. The key to DES is adaptive model aggregation, where the most competent classifiers are selected to make a prediction on a specific test sample. This presentation will discuss the advantages of DES for personalized federated learning and provide comparisons with existing methods.

2 - Generalization of Federated Learning under Different Heterogeneity Levels

Zhenyu Sun, Northwestern University, Evanston, IL, United States

Generalization performance is a key metric in evaluating machine learning models when applied to real-world applications. Good generalization indicates the model can predict unseen data correctly when trained under a limited number of data. Federated learning (FL), which has emerged as a popular distributed learning framework, allows multiple devices or clients to train a shared model without violating privacy requirements. While the existing literature has studied extensively the generalization performances of centralized machine learning algorithms, similar analysis in the federated settings is either absent or with very restrictive assumptions. Here, we aim to analyze the generalization performances of federated learning from a theoretical perspective. We utilize the notion of algorithmic stability to analyze three widely used algorithms, FedAvg, SCAFFOLD, and FedProx, under convex and non-convex loss functions. Our analysis shows that the generalization performances of models trained by these three algorithms are closely related to the heterogeneity of clients' datasets as well as the convergence behaviors of the algorithms.

WB74

Regency - 705

Net-Zero Emissions Energy Systems: Multi-objective Optimization

Invited Session

ENRE: Energy-Climate

Chair: Jacqueline Dowling, Carnegie Science, 260 Panama St., Providence, 94305, United States

Co-Chair: Aleksander Grochowicz, University of Oslo, Oslo, Norway

Co-Chair: Henry Ssembatya, NC State University, RALEIGH, NC, United States

1 - Multi-Objective Optimization for Balancing Techno-Economic and Environmental Considerations in African Hydropower Development

Angelo Carlino, Carnegie Institution for Science, Stanford, CA, United States, Rafael Schmitt, Anna Clark, Andrea Castelletti

In continental Africa, 300 hydropower projects totaling approximately 100 GW of additional capacity have been proposed to meet the growing power demand. Yet, expanding hydropower fragments river reducing the fluvial connectivity and releases greenhouse gas emissions from reservoirs where biomass is digested. Increasingly competitive alternative renewable electricity technologies can be leveraged as alternatives to reduce these impacts with little impact on cost metrics.

Here, we use multi-objective optimization to find Pareto-optimal dam portfolios maximizing hydropower generation while minimizing impacts on river connectivity and emissions from hydropower. We then use these Pareto-optimal solutions to constrain the capacity expansion in a energy system model combining techno-economic and environmental objectives.

Compared to planning for least-cost energy systems, when capacity expansion strategies are designed to balance environmental and techno-economic objectives, electricity prices and total discounted costs increase by at most 1.4% and 0.2%, respectively, while impacts on annual hydropower emissions and river fragmentation are reduced by at least 50%.

These results show that techno-economic analysis can be revised in light of global and local environmental objectives to help policymakers reduce the impacts associated with hydropower development with marginal increases in energy costs.

2 - Targeted power sector decarbonization for improving climate and health benefits of electric vehicles in comparison to efficient gasoline hybrids

Madalsa Singh, Stanford University, Stanford, CA, United States, Inês Azevedo

Electric vehicles and efficient gasoline hybrid vehicles that satisfies the most stringent emissions standards (currently Tier-3 in the US) have emerged as two possible technologies to reduce greenhouse gas emissions and air pollution. Using a spatially resolved emissions inventory, air quality modeling, and temperature and drive-cycle dependent vehicle operations data, we compare electric and Tier-3 vehicles for their climate (greenhouse gas emissions) and health ($PM_{2.5}$ emissions) benefits. In majority of the United States, particularly in metropolitan areas, electric vehicles have higher climate and health benefits that Tier-3 vehicles. However, in certain geographies, Tier-3 hybrids fare better. In these regions, targeted power sector interventions can be deployed to maximize the environmental benefits of transport electrification. We identify 50 high emitting power plants that can be retired as well as regions where natural gas as marginal generator would be sufficient for electric vehicles to have higher climate and health benefits than Tier-3 gas vehicles. Using a modeling approach that combines lifecycle accounting and air quality analysis helps identify solutions that provide climate and health co-benefits.

3 - Land costs impact solar technology choices and carbon mitigation in China

Shi Chen, Carnegie Science, Stanford, CA, United States, Xi Lu, Jiming Hao, Edgar Virgüez, Ken Caldeira, Steven Davis

China has been at the forefront in recent solar photovoltaics deployment. However, only 12% of its utility-scale solar farms use tracking technology, versus 80% in the US. Here, we use a spatially-explicit model of solar electricity costs to explore this disparity. We find that single-axis photovoltaics generate ~20% more electricity than fixed-tilt systems in China, but single-axis systems require 118% more land (national averages). Consequently, electricity from single-axis systems is more expensive when land costs exceed CN¥10.2/m^2 ($1.48/$m^2$, with current average land costs in China at CN¥19.3/m^2 ($2.79/$m^2$)). The decision to deploy fixed-tilt saves land costs but lowers electricity generation per panel, leading to lower CO_2 mitigation- 0.23 tCO_2 less per additional square meter required by tracking. This suggests that countervailing land policies could avoid CO_2 emissions at CN¥166/ton ($24/ton). Our findings may thus have important implications for land-use and energy policy, quantifying fundamental trade-offs related to different solar technologies.

WB75

Regency - 706

Energy Justice I: Building Back Better, Perspectives on Energy Equity in the Just Transition

Invited Session

ENRE: Other Energy

Chair: Kester Wade, Carnegie Mellon University, Pittsburgh, 15232

Co-Chair: Jordan Joseph, Carnegie Mellon University, Pittsburgh, PA, 15217, United States

Co-Chair: Destenie Nock, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Rebates and Grid Decarbonization from the Inflation Reduction Act Promote Equitable Adoption of Energy Efficiency Retrofits

Jordan Joseph, Carnegie Mellon University, Pittsburgh, PA, United States

Residential and commercial buildings consume 75% of the electricity and 40% of the total energy used in the United States, an annual expense of over $400 billion. Furthermore, household energy decisions (personal transportation, space heating, water heating, and other end-uses) account for over 40% of U.S. energy-related emissions which has made residential decarbonization a priority research area. The ability to shift these energy services from fossil fuel to clean electricity has made electrification and other residential energy efficiency retrofits a viable decarbonization pathway. Around half of U.S. homes were built before most building codes and are often occupied by low-to-moderate-income (LMI) families. Equitable electrification provides a unique opportunity to significantly reduce emissions in communities where energy bill savings have the most impact on household finances. We developed the Tradeoff Analysis of Residential retrofits for Energy equity (TARE) model to address the need for holistic approaches that consider unique housing characteristics, the intensity of existing regional grid emissions and fuel prices, health outcomes, and socioeconomic factors. We estimate national baseline and post-retrofit consumption and use an adoption rate metric that considers both public (i.e., climate and health-related emissions impacts) and private (i.e., lifecycle costs and savings) benefits and costs associated with energy efficiency retrofits for space heating, water heating, cooking, and clothes drying, using the National Renewable Energy Laboratory’s End-Use Savings Shapes database.

2 - Heat, Cold, and Costs: Evaluating equity and resilience in household energy use

Kester Wade, Carnegie Mellon University, Pittsburgh, PA, United States

Energy poverty involves excessive spending on energy bills, disconnections, and sacrificing essential services like heating and cooling to limit costs. A 2022 study found 16% of households experience energy poverty due to disproportionate spending. Despite the importance of energy efficiency and conservation for climate goals, barriers especially leave low-income households behind. However, there's a lack of understanding regarding how energy efficiency programs impact different demographics, such as income levels. This study considers ~8,000 households in the mid-Atlantic region and examines the impacts of energy assistance programs on household heating and cooling behaviors over a 3-year period. We focus on the outdoor temperatures at which households turn on heating or air conditioning systems to understand whether low-income households tolerate extreme indoor temperatures to reduce financial stress. Regression analysis helps us estimate the relationship between income and energy use spikes during heatwaves or extreme cold. Energy assistance programs have massive potential to improve physical and financial well-being. This especially affects households that might engage in risky behaviors to mitigate their energy needs. We advance the scientific understanding of whether, why, and how different energy, assistance, efficiency, and conservation programs work in different households. Our framework can help evaluate and refine more than 900 similar programs offered by U.S. utilities.

3 - Near-Term Heat Pump Penetration Given Grid Constraints Under Different Scenarios

Arnav Gautam, Carnegie Mellon University, Pittsburgh, PA, United States

Heating, Ventilation, and Air Conditioning (HVAC) systems (particularly in the residential sector) are getting more critical for human well-being as climate change continues, and in their current form are substantial contributors of CO2 emissions in the United States. Electrifying residential HVAC will be an important part of decarbonization and protecting people from temperature-related health risks. HVAC electrification will occur alongside electrifying other residential energy end-uses, and within established but evolving electrical distribution infrastructure. We perform an upper bound analysis to find the maximum penetration of electrified residential heating and cooling systems that can be expected across the U.S., given existing distribution infrastructure and household electrical capabilities. To answer this, we first alter and run the NREL ResStock model to quantify the impacts of realistic electrified HVAC technologies when installed in residential buildings of different types across different climate zones. This analysis yields insight into 1) how technologies may operate in specific locations, and 2) the electrical load profiles of residential buildings outfitted with them. ResStock outputs inform our estimate of region-specific electrified HVAC technology adoption, which we get by running the Tradeoff Analysis of Residential Retrofits for Energy Equity (TARE) model. We then test the distribution grid's ability to handle the changed electricity load caused by these region-specific adoption targets, using the PNNL Gridlab-D model for power system analysis. The result of this work is region-specific insight into which heat pump technologies may realistically be adopted in the near term.

WB76

Regency - 707

Online and Adaptive Optimization

Invited Session

Computing Society

Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

1 - Learning to Cover: Online Learning and Optimization with Irreversible Decisions

Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States, Michael Li

We define an online learning and optimization problem that aims to achieve a target coverage within a finite horizon. At each period, a decision-maker opens facilities, receives information on the success of each one, and updates a machine learning model to guide future decisions in order to minimize costs under a chance constraint reflecting target coverage. Irreversible facility location decisions give rise to an exploration-exploitation trade-off with commitment. We derive an optimal algorithm and a tight lower bound in an asymptotic regime characterized by a large target number of facilities but a finite horizon. We find that the regret grows at a sub-linear rate that converges exponentially as the horizon becomes increasingly long. This result is robust to the learning environment (with offline data, imperfect learning, and continuous uncertainty), and to the optimization environment (with heterogeneous facilities characterized via a facility-customer graph). These results underscore the benefits of even limited learning, in that even a few rounds of online learning and optimization can provide significant benefits as compared to a no-learning baseline. We conclude with a case study in sustainable infrastructure planning using real-world data to show the benefits of the proposed online learning and optimization approach in practice.

2 - Speed up the Cold-Start Learning in Two-Sided Bandits with Many Arms

Wanning Chen, University of Washington, Seattle, WA, United States, Mohsen Bayati, Junyu Cao

Multi-armed bandit (MAB) algorithms are efficient approaches to reduce the opportunity cost of online experimentation and are used by companies to find the best product from periodically refreshed product catalogs. However, these algorithms face the so-called cold-start at the onset of the experiment due to a lack of knowledge of customer preferences for new products, requiring an initial data collection phase known as the burn-in period. During this period, MAB algorithms operate like randomized experiments, incurring large burn-in costs which scale with the large number of products. We attempt to reduce the burn-in by identifying that many products can be cast into two-sided products, and then naturally model the rewards of the products with a matrix, whose rows and columns represent the two sides respectively. Next, we design two-phase bandit algorithms that first use subsampling and low-rank matrix estimation to obtain a substantially smaller targeted set of products and then apply a UCB procedure on the target products to find the best one. We theoretically show that the proposed algorithms lower costs and expedite the experiment in cases when there is limited experimentation time along with a large product set. Our analysis also reveals three regimes of long, short, and ultra-short horizon experiments, depending on dimensions of the matrix. Empirical evidence from both synthetic data and a real-world dataset on music streaming services validates this superior performance.

3 - Semi-Bandit Learning for Monotone Stochastic Optimization

Viswanath Nagarajan, University of Michigan, Ann Arbor, MI, United States, Arpit Agarwal, Rohan Ghuge

Stochastic optimization is a widely used approach for optimization under uncertainty, where uncertain input parameters are modeled by random variables. Exact or approximation algorithms have been obtained for several fundamental problems in this area. However, a significant limitation of this approach is that it requires full knowledge of the underlying probability distributions. Can we still get good (approximation) algorithms if these distributions are unknown, and the algorithm needs to learn them through repeated interactions? In this paper, we resolve this question for a large class of ``monotone" stochastic problems, by providing a generic online learning algorithm with $\sqrt{(T \log T)}$ regret relative to the best approximation algorithm (under known distributions). Importantly, our online algorithm works in a semi-bandit setting, where in each period, the algorithm only observes samples from the random variables that were actually probed. Our framework applies to several fundamental problems in stochastic optimization such as prophet inequality, Pandora's box, stochastic knapsack, stochastic matchings and stochastic submodular optimization.

4 - Data-Driven Robust Sequential Search

Cagin Uru, Duke University, Durham, NC, United States, David Brown

Models of sequential search arise in a variety of problems, including organizations hiring new employees, firms developing novel technologies, and individuals seeking housing or other investment opportunities. The classical papers in the sequential search literature rely on the assumption that the decision maker (DM) fully knows (or has some prior belief about) the underlying distribution of alternative values. In many applications, however, the DM explores alternatives with limited (if any) prior information on the alternative value distribution. For example, we can consider a seller searching for the best price among several buyers without any access to the distribution of buyers' valuation or any historical data for guidance. Motivated by this consideration, we study a generalization of the classical Pandora's problem in which the alternative value distribution is unknown prior to the search. In each period, the DM can either explore an alternative at a cost to continue the search or select and acquire any previously explored alternative to end the search. The goal is to find a stopping rule that maximizes the worst-case ratio of the expected reward compared with an oracle with full knowledge of the alternative value distribution. We consider several variations of this problem. For each variation, we design simple policies that perform competitively with the oracle policy.

Furthermore, we develop upper bounds on the relative performance of any feasible policy with respect to the oracle benchmark. The resulting upper bounds show that our policies provide nearly optimal performance.

WB77

Regency - 708

Optimization Modeling Software I

Invited Session

Computing Society

Chair: Susanne Heipcke, FICO, Xpress Optimization, Birmingham, United Kingdom

Co-Chair: Bob Fourer, AMPL Optimization Inc., Evanston, IL, 60201, United States

1 - Gamspy: a Pipeline Friendly General Algebraic Modeling Language

Steven Dirkse, GAMS Development Corporation, Fairfax, VA, United States, Adam Christensen, Muhammet Soyturk, Michael Bussieck

Algebraic modeling languages (AMLs) such as GAMS, AMPL and more recently Julia/JuMP have been a cornerstone in the fields of optimization and economics. These tools are popular because they are able to effortlessly link the worlds of algebra and computer science -- that is, the syntax of the AML closely approximates that of handwritten algebra but its execution is automated and scalable. This pedagogical link supports the developer during all phases of software development from original model formulation to performance debugging and final output. The syntax is also important when transferring modeling knowledge between parties (i.e., external code review, developer transitions, etc.). Recent years have seen sophisticated computing tools enter the mainstream. Open source software languages/packages such as Anaconda, Python, Numpy, and Pandas make it easy to work with large data structures efficiently in deployable environments. All of these compute resources mean that data assets are arriving at optimization model instances from an ever diversifying number of start points. All of this creative data pipelining requires new and creative technical solutions from AML developers. In this work we present a Python package called GAMSPy that leverages the flexible nature of Python to construct a true Python-AML. The GAMSPy design philosophy rests on the original benefits of AML languages, specifically that the syntax mirrors algebra. Its set-driven nature, a common element among AMLs, is accomplished with extensive Python operator overloading. Execution of the algebra relies on a GAMS subsystem to avoid the performance and pedagogical pitfalls of object-oriented frameworks.

2 - Modeling Consensus Admm for Progressive Hedging

David Woodruff, University of California-Davis, Davis, CA, United States

A recent paper in Operations Research by Eckstein et al demonstrated that the Progressive Hedging (PH) algorithm can be developed from the same operator splitting framework as the alternating direction method of multipliers (ADMM). In this talk, we look at software constructs to describe the model and pass it to a PH algorithm.

3 - From Algorithms to Applications: Developing and Deploying Analytical and Optimization Models with Fico Xpress Insight

Bruno Vieira, FICO, Spain, Spain

Xpress Insight provides a flexible environment for creating powerful business applications with user interfaces that facilitate the utilization of advanced analytical solutions implemented in Python or Xpress Mosel. The Insight application interface is designed for both technical and non-technical users, enabling experimentation by running several scenarios and comparing outcomes to understand sensitivities implicit to an optimization model. In this talk, we will demonstrate how an Xpress Insight application can seamlessly be built from a standard Python or Mosel optimization model with intuitive drag and drop of front-end visualization components, including multi-scenario charts and tables. Furthermore, we will demo the latest advancements for interacting with the solver during an optimization run via the new Custom Progress Reporting functionalities.

4 - Ampl: Advances in Python Integration, Cloud Deployment, and Generative AI

Filipe Brandão, AMPL Optimization Inc., Mountain View, CA, United States

Python and its vast ecosystem are great for data pre-processing, solution analysis, and visualization, but Python's design as a general-purpose programming language makes it less than ideal for expressing the complex optimization problems typical of prescriptive analytics. AMPL is a declarative language that is designed for describing optimization problems and that integrates naturally with Python.

In this presentation, you'll learn how the combination of AMPL modeling with Python environments and tools has made optimization software more natural to use, faster to run, and easier to integrate with enterprise systems. We will show how AMPL and Python work together in a range of contexts:

- Installing AMPL and solvers as Python packages anywhere

- Fast data transfer from/to Python data structures such as Pandas and Polars dataframes

- Deploying models to the cloud quickly and easily

You'll also see how Generative AI technology is enabling a rapid development process for both AMPL and Python, reducing the time and effort to produce a working application that's ready for end-users.

WB78

Regency - 709

Optimization Algorithms and Applications in Manufacturing

Contributed Session

Chair: Megan Peitz, Numerious Inc., 8139 Brophy Rd, HOWELL, MI, 48855, United States

1 - Variable aggregation for pre-processing non-linear programs

Sakshi Naik, Carnegie Mellon University, Pittsburgh, PA, United States, Robert Parker, Russell Bent, Lorenz Biegler

Large scale nonlinear optimization problems are frequently encountered in several applications such as energy systems, process optimization, etc. Reliable online optimization of these systems is crucial for continuous operation under different conditions. Optimization of these problems using state of the art non-linear solvers can face difficulties in convergence for certain challenging applications. It is therefore, important to develop solution strategies that can reliably solve these systems and are efficient.

In this work we focus on pre-processing of non-linear programs using variable aggregation for better convergence reliability. Variable aggregation involves replacing variables with their constraint expressions which results in a reduced order model with fewer variables and constraints. Variable aggregation has played a crucial role in convergence and efficiency of linear and mixed integer optimization solvers. In NLPs, it has been shown that eliminating equality constraints that define variables can help in convergence[1]. In this talk, we present a new approximate-maximum algorithm for variable aggregation. Problem size, solve-times and problem densities are compared across different aggregation strategies. We demonstrate that variable aggregation in general and particularly the approximate-maximum strategy has the potential to increase convergence reliability in challenging nonlinear programs.

In this talk, we describe our variable aggregation framework developed in Pyomo, give details on the approximate-maximum aggregation and demonstrate that variable aggregation leads to better convergence reliability on various test problems.

1. Parker, R., Nicholson, B., Siirola, J., Laird, C. & Biegler, L. An implicit function formulation for optimization of discretized index-1 differential algebraic systems. *Computers & Chemical Engineering* **168**, 108042 (2022).

2 - Integrated Learning and Optimization for Optimal Economic Dispatch of Power Generation Sources

Imran Pervez, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, Ricardo M. Pinto De Lima, Omar M. Knio

In this work, we propose to use integrated learning and optimization (ILO) to solve economic dispatch (ED) problems. The ED is an optimization problem in power systems solved to obtain power dispatch decisions of electric generators to supply a consumer power demand at minimum operating cost. We consider that the power demand in the ED formulation is unknown and learned using learning algorithms. In the literature, ED has been solved mainly using sequential learning and optimization (SLO), which focuses on accurately training the power demand forecast and then solving the ED problem. However, an accurately trained power demand forecast does not guarantee that the optimal ED decisions are feasible or optimal for the actual demand. To minimize this deviation and the cost of dispatch, an ILO approach is used. Unlike SLO, ILO trains the power demand by minimizing a regret function, designed to minimize the penalties due to supply and demand mismatch, which is evaluated after solving the ED problem within the training phase. The regret function and its gradient are relevant components of this approach, and their derivation is described in detail. After the training phase, the ED problem is solved and the regret function value is consequently determined. The results show that ILO leads to lower regret functions than SLO, thus illustrating its efficacy in minimizing penalties due to supply and demand mismatch.

3 - Building Designs for Individual-Level Estimation: Considerations, Implications and New Tools for Choice-Based Conjoint

Megan Peitz, Numerious Inc, Howell, MI, United States, Trevor Olsen

Choice-based conjoint (CBC) experiments are widely used to understand consumer preferences and willingness to pay for different product features. One important consideration in designing CBC experiments is the balance of attribute levels across the design. This strategy seeks to give every level an equal chance to influence the respondent’s decision in the conjoint design and can work in the majority of cases. However, this work revisits the work of Huber and Zwerina (1996) to determine if utility balanced designs, a design strategy that trades off on level balance while optimizing which alternatives are paired against each other within tasks, could result in better predictions at the individual level. This presentation sets out to explore several different methods of optimizing designs and offers an open-source solution, built in Julia by the Numerious team, to leverage these different design strategies in the future.

The results from this paper show that utility balanced designs perform well in predicting data from both utility balanced and non-utility balanced designs, and that respondents do not seem to be fatigued by utility balanced designs. This would suggest that utility balanced designs could be a successful strategy depending on the attributes and levels being tested. However, we must caution the user of utility balanced designs as some design strategies may result in sparse data at the interaction level. We also believe that further research is needed to understand the differences in willingness to pay estimates between utility balanced designs and traditional, level balanced designs.

Wednesday, October 23, 11:05 AM - 11:55 AM

Summit - Ballroom 1

Reimagining a Healthier Future with Sustainable Supply Chains

Plenary/Keynote Session

Plenary

Chair: Polly Mitchell-Guthrie, Kinaxis, Scottsdale, AZ, United States

1 - Reimagining a Healthier Future with Sustainable Supply Chains

Ann Tracy, Colgate-Palmolive, New York, NY, United States

Sustainability is rising as a priority on the corporate agenda, with increasing attention on supply chain, since this is where key contributions of a company's carbon emissions occur including a focus on the upstream supply chain. Colgate-Palmolive is ranked as one of the best supply chains in the world, and that leadership extends to their innovation in sustainability. They developed the world's first recyclable toothpaste tube and then shared the design with competitors in order to increase adoption. Join this plenary to hear how they are leading the way and learn about the real on-the-ground challenges they are tackling.

Wednesday, October 23, 12:15 PM - 1:30 PM

WC01

Summit - 320

Platform-Based Innovation in Service Operations

Invited Session

Service Science

Chair: Ruijie Zhang, ShanghaiTech University, Shanghai, N/A

1 - Does Kol Matter in Live Ecommerce: An Empirical Study on Tiktok

Kai Tian, Singapore Management University, Singapore, Singapore, Ruijie Zhang, Kejia Hu

The utilization of live-streaming advertising on social media platforms has garnered significant popularity across various industries due to its potential to significantly enhance the performance of advertisements. In this paper, we adopt PSM (Propensity Score Matching) and auto-regressive linear models to investigate the impact of targeting audiences of KOL (Key Opinion Leaders) followers on advertising outcomes. Based on marketing theories and PSM criteria, we incorporate matching variables into the PSM and identified three categories of marketing outcomes: traffic, purchase intention, and engagement. Our empirical results indicate that KOL-targeting reduces traffic, enhances purchase intention but has no effect on engagement. Specifically, KOL-based targeting reduces advertisement clicks and live streaming room entrances by 0.552 and 0.490, respectively. Nevertheless, it has effectively augmented slidecart interactions by 0.147 clicks, order creation by 0.027, and order payment by 0.024. To optimize targeting, we conduct a counterfactual analysis, identifying critical parameters for specific outcomes and fine-tuning non-critical parameters to improve overall marketing effectiveness without additional costs. This comprehensive optimization strategy maximizes desired outcomes and enhances marketing performance.

2 - Allow or Forbid Returns: Online Resale of Functional Collectibles with Authentication and Speculation

Baolong Liu, ShanghaiTech University, Shanghai, China, People's Republic of, Ellen Jin Jiang, Hongyang Zhuang, Yanlu Zhao, Shouyang Wang

Functional collectibles, such as co-branded sneakers and limited-edition handbags, possess dual attributes as both collectibles and usable items. Often found in the resale market due to their scarcity, these items can achieve surprisingly high price premiums. Resale platforms, aiming to prevent counterfeits, offer authentication services before shipping ordered items to consumers. However, this process introduces delays and the risk of price drops while buyers await authentication and delivery. Meanwhile, major platform StockX in the U.S. forbids buyers from returning products after purchasing. However, others like Poizon in China allow returns to protect buyers' rights and control speculation, leading to potential revenue loss for platforms. Motivated by thes observations between platforms in the two countries, this paper investigates the speculation-controlling effects of the return policy and explores operational innovations to enhance platform revenue. We analyze resale data of Yeezy Boost 350 V2 sneakers to understand the economic, behavioral, cultural, and operational factors influencing price premiums. Subsequently, we develop queueing game models based on key operational strategies, namely the matched/returned model and the speculation/collection model. Our findings reveal that while a return policy decreases platform revenue, it effectively controls speculation and stabilizes buyer welfare in schemes with a low cultural intensity towards certain functional collectibles. Conversely, in high cultural intensity schemes, speculation can be self-discouraged. Additionally, to enhance platform revenue, we propose an *order-based prioritization* design that prioritizes authenticating matched orders over returned orders, and we also explore the implementation of *hassle-free after return* strategies.

3 - The Impact of Financial Incentives in Instant Retail: An Empirical Investigation

Ruijie Zhang, ShanghaiTech University, Shanghai, China, People's Republic of

In instant retail, financial incentives are used to address supply-demand imbalances, but their effects in this context are unclear due to platforms bearing the costs and dense store networks. This study investigates the direct and spillover effects of platform-borne financial incentives on order acceptance speed using data from a major Chinese instant retail firm. Regression discontinuity design shows incentives reduce order acceptance time by 23.66%. Network analysis via propensity score matching reveals opposing spillover effects: incentives at focal stores prolong acceptance time at nearby non-incentivized stores but shorten it at incentivized ones. This is explained by siphon and clustering effects. Optimizing incentives considering network effects could reduce costs by 22.69% and total order acceptance duration by 1.55%. Managers should extend incentives to surrounding stores to leverage clustering effects while mitigating siphon effects, and consider store type and distance when designing incentive strategies to enhance efficiency and cost-effectiveness.

4 - Motivating the Drivers with Income-Targeting Behavior: Innovation of Surge Pricing Strategies in Ridesharing Market

Yuelong Zou, Shanghai Jiao Tong University, Shanghai , China, People's Republic of, Xiaowei Mei, Liangfei Qiu, Lai Wei

With the higher level of uncertainties from the supply side in the context of the gig economy, online platforms operating in the ride-hailing business have designed various surge pricing strategies to address the increasing flexibility of the labor force, ensuring a relatively stable level of service. Although the reference-dependent theory has been widely adopted to explain the negativity of short-term wage elasticity, fewer studies have explored how drivers' irrational behaviors, especially the income targeting issue, can reshape the surge pricing mechanisms.

We developed a Stackelberg game model to examine the platform's optimal pricing strategies and the social welfare dynamics, taking into account the drivers' behavioral issues. Our findings indicate that: (1) The optimal solutions can help platforms switch incentive strategies and avoid falling into the circumstance where surge pricing is ineffective due to negative wage elasticity of the labor. (2) The reference-dependent behavior of irrational drivers widens the pricing gaps of both passengers and drivers between peak and off-peak market conditions. In milder market conditions, the platform captures the increased revenue when demand grows, leaving driver wages unchanged. (3) When the driver's income target level is within a certain range, the welfare for the both demand and supply side increases concurrently with the rise in drivers' income target.

WC02

Summit - 321

Decision Making for Advanced Industrial Systems

Invited Session

Quality, Statistics and Reliability

Chair: Congfang Huang, UW Madison, Madison, WI, United States

Co-Chair: Akash Deep, Oklahoma State University, Stillwater, United States

Co-Chair: Jaesung Lee, Texas A&M University, College Station, TX, 77845, United States

1 - Den-Hmm: Deep Emission Network Based Hidden Markov Model with TIME-Evolving Multivariate Observations

Vipul Bansal, University of Wisconsin-Madison, Madison, WI, United States, Shiyu Zhou

A Hidden Markov Model (HMM) is a popular statistical modeling technique for system health state estimation, monitoring, and prognosis. However, most existing HMMs adopted some simple parametric probability distribution as the distribution of observations for a given state and thus cannot capture the intricate dependency of observations on state and possibly other covariates such as time. To address this, we propose a Deep Emission Network-based Hidden Markov Model (DEN-HMM) to capture the complex evolution of multivariate observations with respect to state and time. We also address the challenging issue of state nondiscrimination in DEN-HMM. To overcome this, we propose a regularized loss function that can prevent certain approximate solutions and enhance the state discriminative capabilities of DEN-HMM. The study further demonstrates extensive numerical studies to show the effectiveness of the proposed DEN-HMM, including a case study on steady-state estimation in ultrasonic cavitation-based dispersion processes.

2 - A General Framework for Robust Monitoring of Multivariate Processes

Xiulin Xie, Florida State University, Tallahassee, FL, United States

Statistical process control (SPC) charts provide an important analytic tool for online monitoring of sequential processes. Conventional SPC charts are designed for cases when in-control (IC) process observations are independent and identically distributed at different observation times and the IC process distribution belongs to a parametric (e.g., normal) family. In practice, however, these model assumptions are rarely valid. To address this issue, there have been some existing discussions in the SPC literature for handling cases when the IC process distribution cannot be described well by a parametric form, and some nonparametric SPC charts have been developed based on data ranking and/or data categorization. However, both data ranking and data categorization would lose information in the original process observations. Consequently, the effectiveness of the nonparametric SPC charts would be compromised. In this article, we make another research effort to handle this problem by developing a general process monitoring framework that is robust to the IC process distribution and short-ranged serial correlation. The new method tries to preserve as much information in the original process observations as possible. Instead of using data ranking and/or data categorization, it is based on intensive data pre-processing, including data decorrelation, data transformation, and data integration. Because the distribution of the pre-processed data can be approximated well by a parametric distribution, the design and implementation of the new method is relatively simple. Numerical studies show that it is indeed robust to the IC process distribution and effective for online monitoring of multivariate processes with short-ranged serial correlation.

3 - Vision System Using Mdp-Based Tracking to Create Body-in-White Dimension Control Charts in Automotive Quality

Ramiro Rodriguez Buno, The Ohio State University, Columbus, OH, United States, Yifei Zhang, Yuan Ke Tsai, Chih-Yin Lin

The traditional way to measure gap and flush are still operated manually in an assembly line with specific instrument, such as feeler gauges or dial gauges, which is not easy to obtain continuous data and it requires considerable labor cost and inspection time.

The aim of this project is to develop a vision system capable of automatically and non-invasively measuring dimensions on Body-in-white, including gaps and flushes, thereby reducing the labor costs associated with manual measurement.

The vision measurement system consists of a single camera positioned to face the inspection line. As a vehicle moves through the line, the camera swiftly captures approximately 30 images, ensuring comprehensive coverage of the vehicle's details.

We developed a vision measurement system to measure the dimensions of gaps and flushes for Body- in-white in the assembly line. This system can accurately detect the measurement area and generate dimensional values in real-time with GPU acceleration. We employ Bayesian optimization methods to optimize the parameters of image processing, thereby enhancing the accuracy of edge detection and measure accuracy.

The current detection and tracking method are still based on exhaustive search with a hierarchical level. We are going to model the tracking as a Markov Chain Process to reduce the searching cost by leveraging the historical trajectory.

To predict the gap and flush, a promising direction for future exploration is the application of sequence models like RNN (or LSTM) or Transformers.

4 - Sample-averaged variance-constrained Risk-sensitive Deep RL

Saunak Kumar Panda, University of Houston, Houston, TX, United States, Yisha Xiang

While deep reinforcement learning (DRL) has shown remarkable success in solving complex decision-making problems. However, existing DRL algorithms often lack explicit risk sensitivity mechanisms, due to their focus on maximizing the expected total return value. This has led to suboptimal performance in environments where risk management is crucial. While risk-sensitive RL methods have mostly been limited to tabular and linear methods, risk-neutral RL methods have exploited DRL infrastructure. In this work, we make an attempt to study risk-sensitive DRL with the variance risk criteria involving a sample-averaged version of the total return. Our focus is to find a policy that maximizes the sample-averaged version of total return, subject to a variance constraint bounded by some threshold. We transform the above problem to an unconstrained policy optimization problem and propose an actor-critic algorithm that efficiently updates the policy and other parameters. Finally, we provide numerical studies of our proposed algorithm on several gym environments.

5 - Knowledge Constrained Deep Clustering for Melt Pool Anomaly Detection in Laser Powder Bed Fusion

Erfan Ziad, Arizona State University, Tempe, AZ, United States, Feng Ju, Zhuo Yang, Yan Lu

The rapid expansion of the manufacturing sector has brought laser-based metal additive manufacturing, like laser powder bed fusion, to the forefront of innovation. Yet, its widespread acceptance hinges on overcoming numerous obstacles, including uncertainties regarding part quality when employing standardized materials in additive manufacturing procedures. Clustering techniques are essential in uncovering patterns within data sets, particularly in the field of additive manufacturing, where understanding the behavior of meltpool images is crucial for process optimization. Traditional hierarchical clustering methods often lack the ability to incorporate domain-specific knowledge, limiting their effectiveness in this field. In this study, we propose a novel approach that integrates prior knowledge-constrained hierarchical clustering with encoded meltpool image sequences. By incorporating domain-specific constraints, our approach aims to enhance clustering accuracy and provide more interpretable cluster assignments. Our approach improves clustering performance and provides insights into melt pool image sequences, enabling us to evaluate printed parts' quality.

WC03

Summit - 322

Domain-Informed Machine Learning for Smart Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Weizhi Lin, University of Southern California, Los Angeles, CA, United States

Co-Chair: Qiang Huang, University of Southern California, Los Angeles, CA, United States

1 - Domain-Informed Automated Surface Patch Extraction for 3D Printing Qualification

Weizhi Lin, University of Southern California, Los Angeles, CA, United States, Qiang Huang

In the qualification of shape quality of three-dimensional (3D) freeform products, actual products are compared to their intended designs to ensure geometric accuracy, fit, and functionality. To efficiently quantify shape differences of complex designs, surface patches are typically prespecified and used as the fundamental elements for comparison. However, as manufacturing shifts from mass production to personalization, the increasing variety of designs makes predetermining an infinite variety of surface patches impractical. Additionally, extracting these patches often relies on potentially unavailable prior knowledge, such as the number of surface patches. Moreover, verifying the validity of extracted surface patches can be challenging for unknown types. This work proposes a dimension-reduction framework that includes domain-informed representation, automated extraction, and adaptive verification of surface patches. We identify surface patches as finite primitive elements capable of reconstructing a shape’s deviation patterns. These patches capture distinct deviation patterns in local regions, determined by changes in geometric attributes rather than the original design. We characterize these changes through their smoothness, measured using the Laplacian-Beltrami (LB) operator, and through patterns of change identified by various types of critical points. To automatically extract these patches, we identify critical points as patch centroids through active landmark selection. Expanding from these centroids, we automatically define boundaries to ensure homogeneous LB operators via changepoint detection. We then design verification processes to assess the identifiability of deviation patterns across various surface patches. This framework is demonstrated through case studies in 3D printing.

2 - Physics-Guided Reinforcement Learning for Automatic Fixture Layout Design in Manufacturing Systems

Yinan Wang, Rensselaer Polytechnic Institute, Troy, NY, United States, Tim Lutz, Juan Du, Xiaowei Yue

Fixture layout design is a critical problem in the assembly process of large-scale sheet metal or composite parts. It directly influences the shape deformation of each individual sheet and the quality of the assembled product. The existing research works mainly follow the scheme of developing mathematical-optimization (MO)-based methods to generate the optimal fixture layout via interaction with physical models, finite element analysis (FEA)-based simulations, or its surrogate models. The limitations of MO-based methods can be summarized as that they are memoryless and lack scalability. Memoryless indicates that the experience in designing the fixture layout for one part is usually not transferable to others. Scalability becomes an issue for MO-based methods when the design space of fixtures is large. Furthermore, the usage of surrogate models might also hinder scalability as they might have variations in prediction accuracy when modeling high-fidelity simulations. To address these limitations, we propose a learning-based framework, SmartFixture, to design the fixture layout by training a Reinforcement learning (RL) agent through direct interaction with the FEA-based simulations. The advantages of the proposed framework include: (1) the properties of the learning-based method enable it to generalize to design fixture layouts for unseen scenarios after offline training; (2) the proposed framework has the ability to find the optimal fixture layout over a massive search space; (3) it can be extended to solve other design optimization problems in the manufacturing process. Experiments demonstrate that the proposed framework consistently outperforms the benchmark methods in designing the fixture layouts.

3 - A Personalized Federated Data Analytics Model for Rul Prediction

Yuhan Hu, North Carolina State University, Raleigh, NC, United States, Xiaolei Fang

Federated prognostics enable users to collaboratively develop a failure time prediction model using their combined data while keeping each user's data local and confidential. However, traditional federated models often assume homogeneity in the degradation processes across users, an assumption that may not hold in many industrial settings. To overcome this, our talk introduces a personalized federated prognostic model designed to accommodate users with heterogeneous degradation processes, allowing them to build tailored prognostic models. Numerical studies are presented to demonstrate the effectiveness of our approach.

4 - Privacy-Preserving Process Data Sharing for Metal-Based Additive Manufacturing via Generative Diffusion Modeling

Hugo Rodriguez, Mississippi State University, Starkville, MS, United States, Durant Fullington, Emmanuel Yangue, Chenang Liu, Wenmeng Tian

The objective of this research is to develop a de-identification mechanism for metal-based additive manufacturing (AM) process data, utilizing a generative diffusion modeling framework. The need for AM process data privacy stems from the growth of collaborative data sharing framework, used to develop robust *in-situ* anomaly detection models. However, these data sharing frameworks can compromise the confidential design information within the AM process data. The proposed diffusion modeling approach provides a novel methodology by infusing print path de-identification measures into image augmentation, resulting a surrogate image with design information masked. This allows for more precise data de-identification and the ability to maintain data usability for process anomaly detection, as the surrogate thermal history maintain similarity with the prior data distribution.

WC04

Summit - 323

Emerging Topics in Computational Optimization

Contributed Session

Chair: Jingyu Liu

1 - Binary Factorization of Symmetric Integer Matrices via Integer Programming

Mohsen Mohammadi, Northwestern University, Evanston, IL, United States, Sanjay Mehrotra

Non-negative matrix factorization (NMF) is a fundamental problem in AL/ML, with a wide range of applications across various disciplines. Over the years, numerous variants of NMF have been explored in the literature. Herein, we study the problem of factorizing symmetric integer matrices into binary basis and its transpose. We propose several types of valid inequalities for the problem within its original variable space. We also present a branch and cut scheme to solve the problem, followed by comprehensive computational experiments to showcase the efficiency of our approach.

2 - How Useful Could the Subdominant Eigenvalue be to Estimating Convergence Rate Toward Equilibrium in Queueing Systems

Javad Tavakoli, University of British Columbia-Okanagan, Kelowna, BC, Canada, Winfried Grassmann

For estimating the speed of convergence of stochastic systems toward their equilibrium one often uses the subdominant eigenvalue. In queueing systems, this may not be appropriate. We show this both experimentally through Jensen`s method, as well as theoretically by using diagonalization. We also introduce an interpretation of the diagonalization that is helpful for deciding whether the subdominant eigenvalue should be used. Moreover, we show what model properties lead to complex eigenvalues. This is important because complex eigenvalues complicate the analysis. The theoretical analysis is bolstered by simple examples where the eigenvalues can be expressed analytically, such as the M/M/1/N queue.

3 - Minimax Rate Optimal Algorithms for High-Dimensional Stochastic Linear Bandits

Jingyu Liu, Queen's University, Kingston, ON, Canada, Yanglei Song

We consider the stochastic linear bandit problem with multi-arms over T rounds, where the dimension d of the covariate vectors is potentially larger than T, however each arm-specific parameter vector has at most s non-zero components. First, we study the single-arm case, focusing on the cumulative estimation error of the parameter vector. We show that Lasso has a sub-optimal dependence on d,T in terms of worst-case error in the online estimation setup. Further, we establish the minimax optimality of the OLS post Lasso, which first estimates the support of the parameter vector by thresholding the Lasso estimator and then runs least squares on the selected support. Second, we consider the bandit setup, using the OLS post Lasso as the main estimation method. We propose a three-stage algorithm for arm selection and derive an upper bound of order $s (\log s) (\log d + \log T)$ on its cumulative regret. Finally, we establish a matching lower bound up to a $(\log s)$ multiplicative term, which shows the near minimax optimality of the proposed method. We note that our analysis avoids the restrictive "beta-min" condition.

WC05

Summit - 324

Advanced Methods in Optimization and Prediction

Contributed Session

Chair: Johnson Phosavanh, The University of Sydney, Sydney, 2006, Australia

1 - Guaranteed Bicluster Recovery via Semidefinite Programming

Abiodun Sumonu, The University of Alabama, Tuscaloosa, AL, United States, Brendan Ames

Unlike traditional clustering, in biclustering the goal is grouping both the object and feature sets concurrently into subgroups, called biclusters, based on their expression levels. We consider the densest k-disjoint-biclique problem in which a given weighted complete bipartite graph is partitioned into k disjoint subgraphs such that the sum of their densities is maximized as a model for the biclustering problem. We show that the optimal solution of a semidefinite relaxation of the densest k-disjoint-biclique problem recovers the correct object-feature partitions with high probability whenever the input graph is sampled from a distribution of clusterable bipartite graphs. We discuss necessary and sufficient conditions for exact recovery and empirically investigate how parameters like sparsity, distribution properties, block sizes, and outliers affect recovery through numerical simulations. Our findings highlight theoretical phase transitions, underscoring the existence of critical points for exact recovery of the bicluster structure

2 - Inert Zones in L1 Regression

Botan Citil, University of Nebraska-Lincoln, Lincoln, NE, United States

This work investigates inert zones in online L1 regression, focusing on their role in computational time savings and their relative size with respect to data space. Inert zones, both above and below the regression hyperplane, are regions where additional data points do not affect regression parameters. Characterized as convex, polyhedral shapes, inert zones share the dimensionality of the data set and are defined by intersecting pairs of simplexes with the same barycenter. Our study aims to analyze the computational time savings achieved by identifying inert arrivals and to examine how data set parameters, including cardinality, dimension, and variance, impact these savings and the size of inert zones. Empirical evidence suggests that inert zones can be relatively voluminous, further highlighting their significance in L1 regression. By presenting results on the function and geometry of these novel regions, this paper contributes to a deeper understanding of inert zones in online L1 regression and their practical implications.

3 - Expected Bipartite Matching Distance in A *D*-dimensional L^p Space: Approximate Closed-form Formulas and Applications to Mobility Services

Shiyu Shen, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Yuhui Zhai, Yanfeng Ouyang

The bipartite matching problem has been at the core of many theoretical or practical challenges. Although many well-known algorithms can solve each bipartite matching problem instance efficiently, it remains an open question how one could estimate the expected optimal matching distance for arbitrary numbers of randomly distributed vertices in a D-dimensional L^p space (referred to as a random bipartite matching problem, or RBMP). This paper proposes an analytical model with closed-form formulas (without statistical curve-fitting) that estimate both the probability distribution and expectation of the optimal matching distance of RBMP. Simpler asymptotic approximations of the formulas are also developed for some special cases. These proposed distance estimates could be key for strategic performance evaluation and resource planning in a wide variety of application contexts. To illustrate their usefulness, we focus on mobility service systems where matches must be made between customers and service vehicles that are randomly distributed over time and space. We show how the proposed distance formulas not only provide a theoretical explanation on the conditions under which empirically assumed Cobb-Douglas matching functions can be suitable for taxi systems, but can also be easily incorporated into optimization models to determine taxi operational strategies (e.g., duration of the demand pooling period). A series of simulation experiments are conducted to verify the accuracy and applicability of the proposed formulas under varying numbers of bipartite vertices, varying number of dimensions, and varying L^p distance metrics.

4 - Minimizing the number of late jobs and total late work with step-learning

Johnson Phosavanh, The University of Sydney, Sydney, Australia, Daniel Oron

We study single-machine scheduling problems with step-learning, where an improvement in processing time is experienced if a job is started at, or after, a job-dependent learning-date. We consider minimizing two functions: the number of late jobs and the total late work, and we show that when at least a common due-date or common learning-date is assumed, the problem is NP-hard in the ordinary sense; however, when both are arbitrary, the problem becomes strongly NP-hard. For each problem where at least one of the dates is assumed to be common, we analyze the structure of an optimal job schedule with and without idle time and propose pseudo-polynomial time dynamic programming algorithms. We also show that the problem of minimizing the weighted number of late jobs with step-learning can be solved with a minor change to the algorithms for the unweighted case. In addition, we show that when a common due-date is assumed and no idle time is allowed, the problem of minimizing the total late work is equivalent to minimizing the makespan. Furthermore, we provide a more efficient algorithm to minimize makespan under the assumption of a common learning-date than the one in the existing literature. Lastly, we show that our analysis can also be applied to the case of step-deterioration, where, instead, the processing times of jobs increase at a given date.

WC06

Summit - 325

Strategies in Revenue and Payment Management

Contributed Session

Chair: Darius Walczak, PROS Inc., 3200 Kirby Drive, Suite 600, Houston, TX, 77098, United States

1 - Optimal content-selling strategy with quality-dependent preference consumer

Pengjie Guo, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of

Digital content product is a form of content product widely accepted after the vigorous development of new media, including E-journals, streaming media, digital albums and so on. Such products usually come in the form of content collections, such as multiple articles in an issue of an e-journal or ten songs in an album. They have inherent temporal and experiential features that distinguish them from normal products.

For example, the sequence of content will affect the overall experiential utility of customers and further influence repeat purchase behavior; consumers may be uncertain about the valuation of a content.

An effective way of marketing such a product turns out to be letting consumers to experience a portion of it at a low cost, or even for free, which provide an opportunity for consumers to resolve some (or all) of their uncertainty before buying. Meanwhile, Free trial of partial content products also naturally forms a quality reference point which in turn affects overall customer satisfaction and subsequent repeat purchase behavior.

This literature discusses the case of content products provided by monopoly or duopoly that consumers purchase repeatedly over two periods. We will demonstrate how, depending on these market features, the content provider's optimal sales strategy and intertemporal sequence so as to maximize the total revenue of content provider when consumers are quality reference-dependent.

2 - Buy now pay later: Endogeneity, Heterogeneity, and Intertemporal Preferences in Installment Loan Choices

Jessie Jiang, Cornell University, Ithaca, NY, United States, Christopher Anderson

Buy now pay later (BNPL) is a payment arrangement that enables consumers to make immediate purchases while spreading the payment through fixed monthly installments over time. It has gained rapid adoption across diverse markets including fashion, electronics, and travel. BNPL providers like Afterpay, Affirm and Klarna often present the purchase amount in the form of small monthly payments to attract consumers. We focus on BNPL use in services, where product consumption and payment periods may not be the same – e.g. immediate travel with delayed payment versus (partial) payment in advance of service use. We consider the role of endogeneity, heterogeneity, and intertemporal preferences in BNPL loan choices. Credit score, an essential variable for the loan underwriting process, dictates the loan amount, duration, and interest rates available to the applicants. On the other hand, as credit scores could reflect consumers' risk perceptions and spending habits, they might be also associated with the heterogenous preferences in consumers' loan choices. Additionally, when products are not readily available, such as reserving a trip, consumers would need to balance future consumption with future payments. We introduce three types of temporal elements into the choice model: (1) anticipation utility derived from the total order amount and the time to service use, (2) time discounting for future monthly payments, and (3) present bias owing to the size of down payments. Utilizing a consumer purchase dataset from a BNPL loan provider, we compare a baseline multinomial logit model with models that address endogeneity, heterogeneity and intertemporal preferences.

3 - Huge-Scale Network Revenue Management with Customer Choice: A Primal-Dual Approach

DONGHAO ZHU, The Institute of Statistical Mathematics, Tachikawa, Japan, Hanzhang Qin, Ching-pei Lee, kenji fukumizu

We consider network revenue management with customer choice on a huge scale, which is a central question for many industries including transportation, retailing, and healthcare. The solution to a choice-based linear program usually can be used to guide managers in determining sales volumes to different types of customers to maximize revenue. However, traditional methods typically fail to solve such a linear program at a huge scale due to a lack of tailoring the algorithm to consider the extremely big scale. To address the computational issue, we introduce a first-order primal-dual method SPFOM. Our algorithm requires only a small computational cost in each iteration and enjoys a provably near-optimal convergence rate. This feature is further verified through numerical simulations in comparison to state-of-the-art commercial solvers.

4 - Aspects of Network Revenue Management Control

Darius Walczak, PROS Inc., Houston, TX, United States

Abstract: We analyze several popular control methodologies for network revenue management (network RM) considering their relevance to real-life revenue management and pricing applications. Comparing advantages and disadvantages of both established and novel methods such as network decomposition (resource or product-based), iterative proration or heuristic dynamic programming methods, we point out ways to improve their performance in practice as well as user adoption. Within this context we report on recent machine learning techniques being introduced in the field.

WC10

Summit - 330

Algorithms/Mechanisms for New Priorities in Operations

Invited Session

Revenue Management and Pricing

Chair: Pengyu Qian, Boston University, Boston, United States

Co-Chair: Alys Liang, Stephen M. Ross School of Business, the University of Michigan, Ann Arbor, MI, 48105, United States

1 - Target-Following Online Resource Allocation

Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States, Garrett van Ryzin, Huseyin Topaloglu

We consider a problem in which a decision-maker faces a sequence of requests to be assigned to a set of resources over time. Contrary to classical online resource allocation problems in which there is a fixed budget constraint on the number of resources consumed at the end of the horizon, we assume the time horizon is partitioned into a fixed number of epochs, each associated with a vector of target consumption rates for the resources. The objective is to simultaneously minimize total assignment costs and a convex deviation cost incurred at the end of each epoch which penalizes the difference between the running average consumption of resources and their respective targets. We design a primal-dual algorithm that attains sublinear regret with respect to the hindsight optimal solution, when arrivals are drawn i.i.d. from an unknown distribution. We demonstrate the effectiveness of our algorithm via extensive numerical experiments on a real-world dataset.

2 - Improving Utilization in Deceased Donor Kidney Allocation

Grace Guan, Stanford University, Stanford, CA, United States, Itai Ashlagi, Philipp Strack, Felipe Subiabre

While there are around 100,000 patients waiting for a deceased donor kidney in the U.S., over 25% of recovered kidneys are not placed. We assess the impact of incorporating learning into allocation policies to reduce inefficiencies in the system. Using simulations, we evaluate alternative allocation policies that take advantage of knowing a patient's probability of acceptance against the current sequential offering process. We quantify the reduction in waiting times and increase in utility with alternative allocation policies.

3 - Learning to Rank Under Strategic "Brush Wars"

Yifan Feng, National University of Singapore, Singapore, Singapore, Qinzhen Li, Kevin Chen

We consider a dynamic learning and ranking problem of a digital platform. Uninformed of the products' intrinsic qualities, the platform strives to design a ranking rule that learns from historical traffic data while accounting for sellers' potential manipulation through "brushing" activities, such as fake orders or sales. We study a simple Experiment-Then-Commit (ETC) policy framework and characterize the sellers' strategic responses as a "brushing war" game. We find that in the limit with a long time horizon and many sellers, there is a "self-reinforcing" market equilibrium. That is, the sellers' brushing amounts increase in the product's qualities, thus helping with the platform's complete learning.

4 - Interpretable Data-Driven Policy for Index-Based Yield Protection

Kehan Lu, Duke University, Durham, NC, United States, Jing-Sheng Song, Can Zhang

Index-based yield protection is a novel government subsidy to safeguard smallholder farmers from crop yield loss, utilizing a predefined index such as rainfall rather than the traditional method of actual yield assessments to calculate subsidy payments. In this research, we propose a data-driven framework to determine the optimal subsidy payment amount for a government-issued index-based yield protection policy. Our approach makes minimal adjustments to one widely adopted policy yet significantly improves its effectiveness. It overcomes two main limitations of the existing policy: First, it accounts for the distinct impact of yield prediction errors on risk-averse farmers across different yield levels, enhancing the model's robustness against errors in misspecification. Second, it systematically considers the accuracy of the index when determining the payment to farmers. In addition, we provide high-probability convergence guarantees for the parameters of our algorithm and extend them to performance guarantees in terms of net benefit (i.e., total farmer surplus minus expenditure). Finally, we implement our framework using real-world agricultural, climate, and remote-sensing data to illustrate its advantages.

WC11

Summit - 331

Advancing Online Marketplaces: A Data-Driven Approach

Invited Session

Revenue Management and Pricing

Chair: Yeganeh Alimohammadi, UC Berkeley, Berkeley, CA, United States

1 - Robust Dynamic Staffing with Predictions

Yiding Feng, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Vahideh Manshadi, Rad Niazadeh, Saba Neyshabouri

We consider a natural dynamic staffing problem in which a decision-maker sequentially hires staff over a finite-time horizon to meet an unknown target demand at the end. The decision-maker also receives a sequence of predictions about the demand that become increasingly more accurate over time. Consequently, the decision-maker prefers to delay hiring decisions to avoid overstaffing. However, workers' availability decreases over time, resulting in a fundamental trade-off between securing staff early (thus risking overstaffing) versus hiring later based on more accurate predictions (but risking understaffing). This problem is primarily motivated by the staffing challenges that arise in last-mile delivery operations. A company such as Amazon has access to flexible gig economy workers whose availability decreases closer to the target operating day, but they can be hired at any time before that day if they are available.

We study the above problem when predictions take the form of uncertainty intervals that encompass the true demand. The goal of the decision-maker is to minimize the staffing imbalance cost at the end of the horizon against any sequence of prediction intervals being chosen by an adversary. Our main result is the characterization of a simple and computationally efficient online algorithm that obtains the optimal worst-case imbalance cost, aka., minimax optimal. Our base model considers staffing for one target demand. We also consider generalizations to multiple target demands with either an egalitarian cost objective or a utilitarian cost objective, and to the case where the hiring decisions can be reversed at given discharging costs.

2 - Approximating Optimum Online for Online Capacitated Allocation

Tristan Pollner, Stanford University, Stanford, CA, United States, Alexander Braun, Thomas Kesselheim, Amin Saberi

We study online capacitated resource allocation, a natural generalization of online stochastic max-weight bipartite matching. This problem is motivated by ride-sharing and Internet advertising applications, where online arrivals may have the capacity to serve multiple offline users.

Our main result is a polynomial-time online algorithm which is (0.5+c)-approximate to the optimal online algorithm for c=0.011. This can be contrasted to the (tight) 0.5-competitive algorithms to the optimum offline benchmark from the prophet inequality literature. Optimum online is a recently popular benchmark for online Bayesian problems which can use unbounded computation, but not "prophetic" knowledge of future inputs.

Our algorithm (which also works for the case of stochastic rewards) rounds a generalized LP relaxation from the unit-capacity case via a two-proposal algorithm, as in previous works in the online matching literature. A key technical challenge in deriving our guarantee is bounding the positive correlation among users introduced when rounding our LP relaxation online. Unlike in the case of unit capacities, this positive correlation is unavoidable for guarantees beyond 0.5. Conceptually, our results show that the study of optimum online as a benchmark can reveal problem-specific insights that are irrelevant to competitive analysis.

3 - Approximating Maximum Matching Requires Almost Quadratic TIME

Mohammad Roghani, Stanford University, Stanford, CA, United States, Soheil Behnezhad, Aviad Rubinstein

We study algorithms for estimating the size of maximum matching. This problem has been subject to extensive research. For n-vertex graphs, Bhattacharya, Kiss, and Saranurak [FOCS'23] (BKS) showed that an estimate that is within ϵn of the optimal solution can be achieved in $n^{2-\Omega_\epsilon(1)}$ time, where n is the number of vertices. While this is subquadratic in n for any fixed $\epsilon > 0$, it gets closer and closer to the trivial $\Theta(n^2)$ time algorithm that reads the entire input as ϵ is made smaller and smaller.

In this work, we close this gap and show that the algorithm of BKS is close to optimal. In particular, we prove that for any fixed $\delta > 0$, there is another fixed $\epsilon = \epsilon(\delta) > 0$ such that estimating the size of maximum matching within an additive error of ϵn requires $\Omega(n^{2-\delta})$ time in the adjacency list model.

4 - Online Advertising Strategy for Long-Term Good via IV-Q-Learning with Noisy Instruments

Yueyang Zhong, London Business School, London, United Kingdom

This paper examines how an online advertising system can optimize ad placement to maximize long-term expected revenue, when observed feedback is biased. We develop an advertising strategy by integrating robust causal inference with reinforcement learning. Our proposed strategy is tested in the auction simulator of our partner firm, showing a significant increase in total revenue over the test period compared to baseline strategies.

WC12

Summit - 332

Optimization of Storage and Warehouse Operations

Contributed Session

Contributed

Chair: Marc Goessling, Afresh Technologies, Jersey City, NJ, United States

1 - Path optimization of multiple robots in robotic mobile fulfillment system

ahmad kokhahi, Clemson university, Central, SC, United States, Mary Beth Kurz

The emergence of robotic technology has led to increased attention on automated warehouses, particularly robotic mobile fulfillment systems (RMFS). These systems leverage robots for their efficiency and speed, often incorporating Automated Guided Vehicles (AGVs) to enhance integration. However, a notable challenge in robotic deployment is the potential for collisions, necessitating priority rules for collision management. While existing priority rules, such as task priority or prioritizing based on starting points, have demonstrated efficacy, they do not account for robot downtime and overall system running time. Addressing this gap, this study introduces a novel priority rule, Minimizing Wasting Time (MWT), which prioritizes based on total time wasted due to collisions. Furthermore, in cases where collisions significantly impact vehicle efficiency, task rescheduling is proposed to mitigate collisions and enhance system efficiency. This approach aims to optimize warehouse operations by minimizing downtime and improving overall productivity.

2 - Efficient Selection of Inventory Counts for Perishable Items

Marc Goessling, Afresh Technologies, Jersey City, NJ, United States, Clement Micol, Emily Meigs, Aaron Stern

Automatic ordering systems for perishable inventory optimize the tradeoff between over-ordering and under-ordering. Such systems can only be effective if they have access to accurate inventory estimates. However, frequent inventory counting is expensive in terms of labor cost. The information gain from a manual count varies depending on the accumulated uncertainty in the inventory estimate, which itself depends on factors like shelf life of the item, and reliability of corresponding shipment and sales data. Inventory targeting algorithms aim to find the right balance between information gain and labor cost.

In this talk we propose a targeting solution to select which grocery items are to be counted by a store associate. We implemented a tunable algorithm that can be controlled by setting a few intuitive parameters, including daily counting limits, counting frequency caps and, critically, a lever to trade off effort vs information gain.

In order to optimize these hyper-parameters based on business requirements, we built a backtesting framework that supports offline evaluation of candidate targeting algorithms with different settings. We share practical lessons learned after launching our solution and describe the adjustments we’ve made to the algorithm in response. Finally, we discuss the challenge of measuring inventory estimate accuracy in a system where targeting itself is dependent on inventory accuracy. To this end, we propose a random sampling extension, applicable to any targeting algorithm, that yields unbiased accuracy readings.

WC13

Summit - 333

Blockchain Innovations in Supply Chain Management

Contributed Session

Chair: Yongle Tian, Shanghai Jiao Tong University, Shanghai, N/A

1 - The effect of blockchain implementation on supply chain conflict

Xiaoqi Lyu, Tianjin University, Tianjin, China, People's Republic of, Min Tian, Baofeng Huo

Existing researches underscore the advantages of blockchain technology in enhancing the efficiency and transparency of supply chain management. This study examines the impact of blockchain implementation on supply chain conflicts, revealing the dark side of blockchain application. By enhancing the orientation of technology for social good and reducing dependence on suppliers, the negative impact of adopting blockchain technology can be alleviated. These insights deepen our understanding of the practical applications and strategic significance of blockchain technology in supply chain management.

2 - Blockchain Adoption in the Low-Carbon Supply Chain: A Comparative Analysis of Different Adoption Modes

Yongle Tian, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Chunguang Bai

Companies can reduce their carbon footprint through both direct emission reduction efforts and indirect offset in emissions trading. Blockchain plays a vital role in addressing the challenges faced by companies in reducing their carbon footprint and has been increasingly adopted across industries. There are mainly three blockchain adoption modes—government adoption, manufacturer adoption, and retailer adoption—yet the optimal mode remains unclear both in practice and in the literature. By fully incorporating blockchain's three benefits (reducing emission reduction cost, reducing transaction cost in emissions trading, and enhancing consumers' green trust) and three costs (setup cost, operational cost, and usage fees), we compare the three blockchain adoption modes in terms of LCSC's sustainability, profitability, and social welfare. Key findings are as follows. First, regarding sustainability, LCSC may rely more on indirect offset after blockchain adoption, highlighting the trade-off between direct and indirect methods of reducing the carbon footprint. Second, regarding profitability, greater efficiency improvement in emission reduction through blockchain always benefits users, whereas it is not true for the investor. Third, retailer adoption is most effective in encouraging LCSC's direct emission reduction efforts and maximizing the manufacturer's profit and social welfare; the retailer's profit will also be highest if the setup cost coefficient and total carbon footprint per unit of product are low. Finally, blockchain's impacts on companies' direct emission reduction efforts and profitability are different across supply chains with different carbon footprint levels. We also provide valuable managerial insights for LCSC and policymakers on driving sustainable development and enhancing social welfare.

WC14

Summit - 334

Technological Advances in E-Commerce and Retail

Contributed Session

Chair: Behrooz Davazdahemami, University of Wisconsin-Whitewater, W164N8925 Water St, Milwaukee, WI, 53051, United States

1 - Understanding the Dynamics of Deal-of-the-Day Platforms: a Multi-Level Analysis

Behrooz Davazdahemami, University of Wisconsin-Whitewater, Whitewater, WI, United States, Elham Rasouli Dezfouli

Daily deal platforms like Groupon have experienced rapid growth in E-commerce over the last decade. Numerous studies have explored what influences consumer purchasing behaviors on these platforms, but they often overlook the impact of deal types (whether they are hedonic or utilitarian). Furthermore, existing research fails to fully account for the effects of time-variant factors such as remaining sales time and previous sales.

In this study, drawing on the theory of duality modes, we examine how deal type moderates the influence of various time-variant and static factors on sales within Deal-of-the-Day (DoD) platforms. Leveraging a substantial dataset from an actual DoD website, we utilize a hybrid within-between panel regression approach to reveal the evolving dynamics of each determinant factor within and across deals over time.

2 - Beyond the Checkout Lines: An Empirical Analysis of Self-Service Technologies in Retail Stores

Jeongha Kim, Korea University, Seoul, Korea, Republic of, Hyeokkoo Eric Kwon, Dongwon Lee, Hyunseok Lee, Kyuhan Lee

Technological progress and changes in consumer habits are significantly reshaping the retail industry, with self-service technologies (SSTs) playing a pivotal role in modernizing the shopping experience. Despite their widespread use, there is still a significant gap in understanding how SSTs impact essential business metrics such as sales and labor productivity, particularly in traditional retail environments. Existing studies on customer segmentation often focus on basic demographics (e.g., age and service usage duration). Against this backdrop, this study examines the effects of SSTs on sales dynamics, labor productivity, and customer segmentation in the retail context. The result of our study based on the difference-in-differences method identifies the reduction in traditional Point of Sale (POS) transactions, suggesting a shift toward fewer but larger purchases. We also observe fewer cashiers and a lower transaction workload, indicating increased efficiency at POS counters. Additionally, there is a notable decrease in unregistered customers after SST implementation, suggesting a shift toward a more identifiable and engaged customer base. This is vital for retailers looking to implement more focused marketing strategies and improve operational efficiency. The research provides detailed insights into how different types of customers and operational timings are affected by SSTs, offering valuable perspectives for future retail strategies and technology integrations. Overall, this study not only expands the existing body of literature on the effects of SST but also guides retailers in optimizing the use of technology to enhance both customer experiences and operational effectiveness in the digitalized retail landscape.

3 - Will Providing Return-Freight-Insurances Do More Good than Harm to Dual-Channel E-Commerce Retailers?

Jiaxin Lin, The University of Hong Kong, Hong Kong, China, People's Republic of, Tsan-Ming Choi, Yong-Hong Kuo

Today, e-commerce retailers commonly operate in a dual-channel mode. Return freight insurance (RFI) is an emerging measure to resolve online shopping disputes with product returns. In practice, some dual-channel e-commerce retailers offer RFI to consumers, while others do not. We build consumer-utility-based analytical models to study the retailer's optimal pricing decisions and values of RFI. In the basic models, we examine three cases, namely Case N (RFIs are not provided), Case R (retailer purchases RFI for consumers), and Case C (consumers pay for RFI). Comparing these three cases, we uncover that the retailer who purchases RFI for consumers does not necessarily charge a higher price. We show that if the RFI premium is sufficiently (moderately) low, it is more beneficial for consumers (the retailer) to pay for the RFI. We analytically prove that (i) when the product's salvage value is polarized or the return freight cost is low, using RFI can help increase consumer surplus (CS), (ii) when the salvage value is sufficiently high, the social welfare (SW) with RFI is higher than the case without RFI. In the extended models, we explore the situation in which consumers can decide whether to purchase RFI as well as the channel to buy the products. We find that (i) the retailer should provide RFI only when the product's cost, salvage value, and return freight cost are all high, and (ii) offering RFI can increase CS but hurt SW. We also consider various extended models to prove the robustness of the research findings.

4 - Modeling the Growth of Unorganized Pharmaceutical Retail Sector: A Critical Review Based on Indian Pharmaceutical Industry

Koushik Mondal, Indian Institute of Technology Kharagpur, Kharagpur, India, Balagopal G Menon

Retail industry serves as the final member of the supply chain that directly faces customer demand fluctuations and, by fulfilling them, generates cash flow which is the lifeline of all supply chain members. Presently, organized, unorganized and e-commerce retailers are fiercely competing for market share in the emerging countries for last two decades. Surprisingly, unorganized retailers have demonstrated remarkable resilience in the face of this competition. The literature on this 'competition' among different retail formats in emerging economies is limited and mostly empirical in nature. Most of the researches are survey or data based and analyzes the retail market dynamics from a model-free perspective. Empirical studies can capture the immediate effect of retail competition when a new organized retailer enters into the market, which initiates dynamic changes in consumer purchasing behavior, infrastructure, and public policy, particularly in emerging economies. It is pertinent to analyze how the retail landscape will become stable in long run and predict the equilibrium state especially in a developing economy like India. Moreover, big retail chains and e-commerce players are offering low prices, customer loyalty programs and extensive product variety that the unorganized sector is struggling to compete with. Presently, the unorganized sector has seen a drastic decline in market share due to high competition from organized and e-commerce, and demonetization; and Indian pharmaceutical industry is no exception to it. In this backdrop, present critical review discusses the existing researches towards understanding the unorganized pharmaceutical retail supply chain dynamics in general and specifically with Indian context.

5 - How Technology-Enabled Price Discrimination Interacts with Upstream Encroachment

Yu Jiang, University of Science and Technology of China, Hefei, China, People's Republic of, Jie Wu, Xiang Ji

With the rapid growth of the platform economy, an increasing number of online platforms and firms are using information technology to price discriminate based on the browsing and purchase histories of consumers. Meanwhile, in addition to selling through the large retail platforms, increasing numbers of firms from a variety of industries are establishing their own direct channels. This study examines the interaction between an upstream firm's encroachment decision and a downstream online platform's technology-enabled price discrimination (TPD) strategy. We find that an upstream firm has less incentive to encroach if the platform commits to adopting TPD than if it does not and prefers to enter if wholesale-TPD is allowed. Our research, thus, is the first to suggest that TPD can critically affect a firm's decision to encroach. On the platform side, we show that, in anticipation of a firm's encroachment, the platform is more likely to use TPD when the technology investment cost is low. We further identify the conditions under which a firm offers direct-TPD or wholesale-TPD. Our results show that a firm prefers to not adopt TPD in the direct channel, and the firm would adopt wholesale-TPD when it is not encroaching.

WC15

Summit - 335

RMP Session

Invited Session

Revenue Management and Pricing

Chair: Neha Sharma, The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104, United States

1 - To commit or not to commit? Analyzing reservation platforms

Neha Sharma, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

Analyzing reservation platforms

2 - Data Sharing and Website Competition: The Role of Dark Patterns

Chiara Farronato, Harvard Business School, Boston, MA, United States

Companies often obfuscate cookie preference policies and consent forms in order to nudge users to share their data more than they would otherwise do. These practices, known as "dark patterns," are pervasive despite their effects being poorly understood. With the use of Webmunk, a browser extension developed for research studies of this type, we explore the effects of dark patterns on consumers' privacy choices. We are particularly interested in exploring the role of limiting and emphasizing options in order to increase data sharing. We find that dark patterns are effective, particularly for less popular websites, for which the baseline willingness to share data is lower than for popular websites. Given that the effects of dark patterns differ across large and small companies, the ensuing firms' competitive advantages in consumer data are affected by regulation forbidding them.

3 - Online Matching with Heterogeneous Supply and Minimum Allocation Guarantees

Raghav Singal, Tuck School of Business at Dartmouth, Hanover, NH, United States, Garud Iyengar

Motivated by our interaction with a labor marketplace, we focus on matching jobs to workers with heterogeneous preferences. In each period, jobs arrive sequentially and need to be matched to a worker in an online manner. Each worker has three parameters: her work quality, capacity, and the minimum number of jobs she desires. The platform wishes to maximize the quality of matches via its online matching policy. We use our model to understand the limitations of simple policies and propose an optimal index-based policy. We supplement our theory with extensive numerics.

4 - Designing Effective Fundraising Campaigns

Pramit Ghosh, The University of Texas at Dallas, Richardson, TX, United States, Ignacio Rios, Anyan Qi

Charitable donations are a vital source of funding for non-profit organizations, enabling them to carry out their mission of addressing social issues and providing support to those in need. To enhance donations, fundraisers employ various designs aimed at increasing donations. However, budgets are often limited, and fundraisers must strategically decide on the optimal design choice that maximizes overall donations. This paper investigates the effectiveness of two design choices in the context of fundraising: (i) the incentive mechanism focusing on matching and gift unlock programs; and (ii) the solicitation mechanism, i.e., whether donations arrive simultaneously or sequentially. Our research findings highlight the importance of design choices in fundraising and provide insights to maximize donation outcomes.

WC16

Summit - 336

Implications of Uncertainties and Market Power in Electricity Market Pricing

Invited Session

Revenue Management and Pricing

Chair: Cheng Guo, Clemson University, Clemson, SC, United States

Co-Chair: Bolun Xu, Columbia University, New York, NY, United States

1 - Endogenous Entry in Networked Markets

Cheng Guo, Clemson University, Clemson, SC, United States, Ozan Candogan, Jiayi Wang

We propose a competitive equilibrium model which allows price to form endogenously on a networked market with profit-maximizing firms. Our model considers a general topology, and studies the effect of various factors on pricing, such as demand levels and network structures. Our findings shed light on market entry incentives.

2 - Economic Capacity Withholding Bounds of Competitive Energy Storage Bidders

Bolun Xu, Columbia University, New York, NY, United States, Xin Qin, Ioannis Lestas

Economic withholding in electricity markets refers to generators bidding higher than their true marginal fuel cost, and is a typical approach to exercising market power. However, existing market designs require storage to design bids strategically based on their own future price predictions, motivating storage to conduct economic withholding without assuming market power. As energy storage takes up more significant roles in wholesale electricity markets, understanding its motivations for economic withholding and the consequent effects on social welfare becomes increasingly vital. This paper derives a theoretical framework to study the economic capacity withholding behavior of storage participating in competitive electricity markets and validate our results in simulations based on the ISO New England system. We demonstrate that storage bids can reach unbounded high levels under conditions where future price predictions show bounded expectations but unbounded deviations. Conversely, in scenarios with peak price limitations, we show the upper bounds of storage bids are grounded in bounded price expectations. Most importantly, we show that storage capacity withholding can potentially lower the overall system cost when price models account for system uncertainties. Our paper reveals energy storage is not a market manipulator but an honest player contributing to the social welfare. It helps electricity market researchers and operators better understand the economic withholding behavior of storage and reform market policies to maximize storage contributing to a cost-efficient decolonization.

3 - Pricing Stochastic Rolling-Window Dispatch

Cong Chen, Cornell University, Ithaca, NY, United States, Lang Tong

Integrating renewables, battery energy storage systems, and electric vehicles into power grids is crucial for decarbonization, yet it poses challenges to real-time grid operations due to uncertainties and intertemporal correlations. Rolling-window dispatch enables the power system operator to centrally schedule generations and demands akin to Model Predictive Control. By anticipating future time horizons, optimizing system dynamics, and accounting for stochastic renewables and demands, rolling-window dispatch proves effective in real-time operation under uncertainty. However, it can lead to a "missing-money problem," wherein market participants may be under-compensated due to imperfect forecasts and limited lookahead horizons.

In this presentation, we will demonstrate the inadequacy of uniform pricing mechanisms in providing dispatch-following incentives in rolling-window dispatch. Consequently, discriminative out-of-market uplifts are required to address the "missing-money problem." To tackle this issue, we propose a non-uniform pricing mechanism that incorporates discriminative intertemporal constraint-induced opportunity costs. Our approach eliminates the missing money problem and ensures dispatch-following incentives independent of forecast accuracy in stochastic rolling-window dispatch.

4 - Prequalifying Residential Distributed Energy Resources (Ders) for Providing Flexibility to Power Markets

Yijiao Wang, Johns Hopkins University, Baltimore, MD, United States, Benjamin Hobbs, Elina Spyrou, Michael Blonsky

With increasing amounts of intermittent renewables in the electricity mix, managing the high variability and uncertainty of these renewables is a growing challenge for grid operators. Flexibility from residential DERs can be used to hedge system reliability risks (supply-demand imbalances). We develop an advanced prequalification and validation technique for demand-based DERs to select aggregated small-scale residential customers for flexibility provision in day-ahead markets based on user characteristics of interest. Specifically, a set of indices depicting the reliability and responsiveness of DERs are developed and demonstrated focusing on the control of HVAC systems and water heaters. Based on the reliability indices, grid operators can identify the most cost-effective small-scale aggregation for various power market applications.

WC17

Summit - 337

Innovative Platform Strategies

Invited Session

Revenue Management and Pricing

Chair: Yao Cui, Cornell University, Ithaca, NY, United States

Co-Chair: Wee Kiat Lee, Nanyang Technological University, Singapore, Singapore

1 - Voice Chatbot Design: Leveraging the Preemptive Prediction Algorithm

Shuai Hao, Southern University of Science and Technology, Shenzhen, China, People's Republic of, Yuqian Xu

This study delves into the design of voice chatbots, focusing specifically on leveraging prediction algorithms to enhance their creativity, and assesses the effects on different business performance metrics.

2 - Freelancers OR Employees? a Field Experiment with a Food Delivery Platform

Wee Kiat Lee, Nanyang Technological University, Singapore, Singapore, Yao Cui, Qi Li

The ongoing debate surrounding the regulation of gig platforms to classify service providers as employees rather than freelancers has generated significant industry discussions. However, the potential impact of such a transition on the operational efficiency of gig platforms is largely unknown. In this paper, we adopt an operational perspective to examine this issue. We collaborate with a food delivery platform and implement a field experiment in which orders are randomly routed to freelancers and employees. Using the experimental results, we compare the job performance between freelancers and employees, and derive insights into how a platform should adjust its operational policies when it changes the recruitment status of service providers.

3 - Information Aggregators to Transaction Facilitators: a Natural Experiment on a Trucking Platform

Vivek Choudhary, NBS, NTU, Singapore, Singapore, Zhenming Shi, Serguei Netessine

The e-commerce platforms can be broadly segmented into two primary types: information aggregators and transaction facilitators. The difference between the two types of platforms lies in integrating a transaction system within the platform. This transaction system allows parties to transact on the platform after being matched, unlike the information aggregators, where the transactions happen outside the platform. Many platforms start as information aggregators, and once they have large volumes of sellers and buyers, they transform into transaction platforms. How does this transformation impact different stakeholders? This study examines this question. Utilizing a quasi-experiment conducted by a major Chinese freight logistic platform, this research employs a Difference-in-Differences (DID) analysis of 2.6 million transportation records over five months in 2023 to empirically assess the impact of integrating transaction systems. Our findings indicate that enabling transaction functionality not only increases the weekly income of drivers by an average of 15% but also reduces the time for order acceptance and delivery costs, thereby benefiting all stakeholders involved. The analysis reveals that the increased income results from enhanced trust and better utilization of resources, where drivers are more willing to accept orders from unfamiliar requesters. However, the rate drivers earn drops, and for the same total amount, the drivers need to drive for longer distances and transport more goods.

WC18

Summit - 338

Novel Applications of Pricing and Demand Modeling

Invited Session

Revenue Management and Pricing

Chair: Pelin Pekgun, Wake Forest University, Winston-Salem, NC, United States

Co-Chair: Sanghoon Cho, Texas Christian University, Fort Worth, TX, United States

1 - Hotel Room Category Overbooking Limits with Paid Upgrade Offers

Andrew Vakhutinsky, Oracle Labs, Burlington, MA, United States, Natalia Kosilova, Farbod Ekbatani

We describe an approach to optimizing booking limits for hotel room categories when the hotel guests are offered to pay for the upgrades to premium rooms. As is often the case, the demand for basic room categories may exceed their physical capacity while demand for premium rooms may be below the category capacity. In this situation, the hotel operators often allow excessive bookings in the basic category followed by complimentary upgrades in the premium category for the customers booking at the basic rate, which results in the loss of potential revenue. We propose a mechanism to couple overbooking of basic categories with offering basic category guests an upgrade to a premium room for a fee similar to the approach used in the Oracle Nor1 product. The optimal solution balances the marginal return rate of the basic category at the given booking limit with the expected future sales of the premium category at a certain level, which allows hotel operators to increase the booking limit and their key performance indicators such as revenue per room. We explain how this approach can be generalized to the case of multi-day stays and how to account for the cancellations of the bookings.

2 - Demand Estimation with Missing Data: Imputation-Based Approach

Sanghoon Cho, Texas Christian University, Fort Worth, TX, United States, Danhyang Lee, Andrew Vakhutinsky

Modeling demand and accurately estimating choice models based on product attributes are crucial for firms to forecast demand and optimize production and pricing strategies. Discrete choice models, such as the multinomial logit model, are popular in industries like retail, airlines, and hotels, especially with the rise of online platforms and increased pricing transparency. However, missing covariates challenge the practical application of these models. This paper addresses the issue of missing covariates in estimating discrete choice models, focusing on scenarios where essential covariate information, such as prices, is missing for unsold products. We propose a novel imputation method based on an odds ratio modeling framework, extending the exponential tilting model to handle multivariate nonignorable missing mechanisms. Additionally, we introduce a simplified dependence structure among covariates to manage severe missing problems.

3 - The Spillover Effects of Giveaways on Gaming Platforms: Evidence from a Quasi-Experiment on Steam

Campbell Clarkson, University of South Carolina, Columbia, SC, United States, Simin Li, Necati Tereyagoglu, Sriram Venkataraman

We investigate the spillover effects of competitor promotions on the Steam platform's user engagement, utilizing the introduction of a game giveaway program by Epic Games, one of Steam's competitors. We also examine whether these spillover effects are driven by increased online popularity of giveaway games using traffic data from Twitch.tv, a livestreaming platform. Finally, we assess which game- and platform-specific features may act as mechanisms for spillovers to occur.

4 - Subscription vs. Selling for Digital Goods Supply Chains in the Presence of Piracy

Hongseok Jang, Tulane University, New Orleans, LA, United States, Janice Carrillo, Kyung Sung Jung, Young Kwark

Recent press coverage of piracy and digital goods touts the practice of subscription as a “piracy killer.” However, the effectiveness of digital goods subscriptions remains controversial in terms of the profitability for different supply chain members. Specifically, the dearth of existing studies concerning business model choices in distribution channel structures indicates that the literature has yet to provide a comprehensive answer to this question. Therefore, we develop an analytical model to investigate the optimal business model choices for digital good firms with the existence of digital piracy in a centralized supply chain (CSC) or a decentralized supply chain (DSC) explicitly considering heterogeneous consumer usage rates. Unique to the current literature, we find that illegal copies serve as a substitute for a different set of consumers, depending on the business model used by the firms. There are circumstances under which the firms optimally allow high-usage consumers to adopt illegal copies in the subscription model. In contrast, illegal copies can also serve as a substitute for the low-usage consumers in the selling-ownership model. Unlike current literature, we identify situations in a CSC whereby the selling-ownership model (a) is more profitable and (b) has fewer illegal goods adopted than the subscription model when piracy is present in the market. When analyzing a DSC, we find that the existence of piracy can actually aid in the coordination of the supply chain because of the interplay between the consumer targeting effect and the inefficiency mitigation effect enabled by digital piracy.

WC19

Summit - 339

Energy Power Systems Management

Contributed Session

Chair: Jehum Cho, Erasmus School of Economics, Rotterdam, N/A, Netherlands

1 - Assessing Zonal System Inconsistencies on Power Network Security in Real-Time Operations

Jehum Cho, Erasmus University Rotterdam, Rotterdam, Netherlands, Anthony Papavasiliou

Zonal market systems encounter numerous challenges associated with network congestion. Controlling cross-border power flows at the zonal level further exacerbates issues such as unexpected violations of capacity limits on both intra- and inter-zonal transmission lines. In day-ahead markets, re-dispatch processes can alleviate some of these problems. However, in balancing markets, there is no time to address these inconsistencies. Such discrepancies in real-time operations are directly related to the safety concerns of the network. In this talk, we introduce a method for analyzing the discrepancy between nodal and zonal systems. Nonlinear behaviors of zonal systems, due to the merit order for each zone, are captured using merit-order constraints with binary variables. This allows us to represent the true feasible region of zonal net positions that ensures security in real-time operations. By comparing this feasible region subject to merit-order constraints with the one derived from existing zonal models, one can determine the extent of system distortion in terms of feasibility. We analyze the conditions under which zonal systems can avoid congestion and evaluate the impact and usefulness of the minimum cross-border capacity policy.

2 - Bidding on Combinatorial Electricity Auctions

Thomas Hübner, ETH Zurich, Zurich, Switzerland, Gabriela Hug

Day-ahead markets are one of the core elements of a liberalized energy system and allow energy to be traded for the next 24 hours in advance. In Europe, it is designed as a combinatorial auction where agents can bid not only on power for single hours but also on bundles of power in different hours by so-called block bids. This allows agents to express their complementarities between hours caused by inter-temporal constraints or quasi-fixed costs. However, it also complicates the bidding decision as the number of bundles feasible to an agent might be exponential regarding the number of hours. As an auctioneer constrains the number of bids by an agent to ensure a tractable market clearing problem, the question arises of which bids to select. In case of a sub-optimal bid selection, agents do not achieve their maximal attainable utility, leading to an overall welfare loss. We take the agent's perspective to study the action design and give decision-support tools to place bids more effectively.

3 - Exploring “Antifragility” for power systems: concepts, applications, and call for actions

Ming Jin, Virginia Tech, Blacksburg, VA, United States

This study investigates the transformative concept of antifragility and explores computational mechanisms to enhance the resilience of power systems. Antifragile systems possess the unique ability to improve and grow stronger when exposed to challenges. To operationalize this concept, we introduce two recent computational approaches for power system resilience: (1) in-context learning with transformer architecture and weak classifiers for zero-day attack detection, and (2) a hierarchical meta-reinforcement learning framework for critical load restoration. The presentation further discusses the importance of a safety-first approach, complementing existing solutions, and leveraging computational innovations to build antifragile power systems. We aim to inspire a collective effort to lay the foundation for computational antifragility, paving the way for more resilient and self-improving infrastructure in the face of increasing uncertainties and threats.

4 - Optimizing Investments in Transmission and Storage with Stochastic Dominance

Miguel Carrion, Universidad de Castilla - La Mancha, Toledo, Spain, Ruth Domínguez, Sebastiano Vitali

Increasing intermittent renewable generation to meet climate goals requires transforming current power systems to ensure rapid and flexible responses to energy flow changes. This work addresses the investment problem faced by transmission expansion planners seeking to expand the transmission network and deploy storage units for decarbonization efforts. We propose a two-stage stochastic problem considering uncertainty in demand growth. Representative days are used to characterize demand and renewable power variability. To optimize expansion strategies while satisfying emission limits, we include second-order stochastic dominance constraints. Numerical analyses using a realistic power system show that installing storage units is vital for efficiently integrating renewable generation. Additionally, investments in the transmission system focus on lower-voltage lines. Formulating the problem with stochastic dominance constraints leads to increased second-stage investments, facilitating better adaptation to evolving demand growth.

WC20

Summit - 340

Applications of Federated and Distributed Learning Frameworks

Invited Session

Decision Analysis Society

Chair: Michael Ibrahim, Georgia Institute of Technology, Atlanta, GA, United States

1 - Federated Anomaly Detection in Vertically Partitioned Datasets

Ayush Mohanty, Georgia Institute of Technology, ATLANTA, GA, United States, Paritosh Ramanan, Dan Li, Nagi Gebraeel

In today's industrial landscape, the proliferation of Internet of Things (IoT) technology has ushered in an era where large-scale systems generate high-dimensional data from diverse sources. These systems are characterized by multiple interconnected components, each equipped with numerous sensors, resulting in a plethora of multivariate time-series data. However, the logistics of centrally aggregating this data pose significant challenges. In this talk, we present a federated learning framework tailored to address the complexities of systems where raw time-series data are distributed across a multitude of components. Unlike traditional methods that require central aggregation of sensor data, our framework operates by capturing the inter-dependencies across components without the need for data movement. We present the theoretical underpinnings of the framework, demonstrating its performance against model-based approaches reliant on centralized data aggregation. We also showcase its practical utility by applying it to the anomaly detection problems, highlighting its ability to assess root causes in real-time, without moving the high-dimensional sensor data.

2 - A Distributed Distributionally Robust Support Vector Machine

Michael Ibrahim, Georgia Institute of Technology, Atlanta, GA, United States

We study the problem of training a distributionally robust (DR) support vector machine (SVM) in a distributed fashion over a network comprised of G clients and one central server, where the clients can only communicate with the server, and the data collected locally at the clients cannot be shared. To achieve this, we propose a novel ambiguity set that relies on local Wasserstein balls centered at the empirical distribution of the data at each client. We then demonstrate that this ambiguity set allows for the DR problem to be reformulated in a separable fashion. We then study out-of-sample performance guarantees associated with our ambiguity set, demonstrating that the probability that the true distribution $\mathbb{P}$ governing the data belongs to the ambiguity set is related to the radii of the local Wasserstein balls. Subsequently, we propose two algorithms that can be used to solve our proposed model: i) a subgradient method-based algorithm, and ii) an alternating direction method of multipliers (ADMM)-based algorithm. We derive the optimization problems to be solved by each client and closed-form expressions for the computations performed by the central server during each iteration for both algorithms. Finally, we carefully examine the performance of the proposed algorithms in a series of thorough numerical experiments utilizing both simulation data and popular real-world datasets.

3 - Federated Multilinear Principal Component Analysis with Applications in Prognostics

Yuqi Su, North Carolina State University, Raleigh, NC, United States, Chengyu Zhou, Tangbin Xia, Xiaolei Fang

Multilinear Principal Component Analysis (MPCA) is a widely utilized method for the dimension reduction of tensor data. However, the integration of MPCA into federated learning remains unexplored in existing research. To tackle this gap, we propose a Federated Multilinear Principal Component Analysis (FMPCA) method, which enables multiple users to collaboratively reduce the dimension of their tensor data while keeping each user's data local and confidential. The proposed FMPCA method is guaranteed to have the same performance as traditional MPCA. An application of the proposed FMPCA in industrial prognostics is also demonstrated. Simulated data and a real-world data set are used to validate the performance of the proposed method.

4 - Distributed Estimation of Inter-Dependencies in Multi-Component Systems

Nazal Mohamed, Georgia Institute of Technology, Atlanta, GA, United States, Ayush Mohanty, Nagi Gebraeel

The integration of sensors and IoT devices has created a network of interdependencies between different systems of an enterprise, each with its own unique data storage and computing resources. This appears in industrial applications like power plant operations, supply chain networks, and some large manufacturing enterprises. Understanding the interdependencies between components of these decentralized systems is important for improved productivity and reliability. In this presentation, we consider a stylized setting where the objective is to estimate a state matrix of a linear time-invariant (LTI) state space model in a decentralized manner without the need for data centralization. The state matrix helps us understand and quantify these interdependencies. This can be very beneficial in settings where system components generate high-dimensional data. We provide theoretical insights into system identifiability in a decentralized setting. We also examine the accuracy of our decentralized estimation approach and study its scalability.

5 - SplitVAEs: Decentralized Scenario Generation from Siloed Data for Stochastic Optimization Problems

H M Mohaimanul Islam, Oklahoma State University, Stillwater, OK, United States

Stochastic optimization problems in large-scale multi-stakeholder networked systems (e.g., power grids and supply chains) rely on data-driven scenarios to encapsulate uncertainties and complex spatiotemporal interdependencies. However, centralized aggregation of stakeholder data is challenging due to privacy, computational, and logistical bottlenecks. In this paper, we present SplitVAEs, a decentralized scenario generation framework that leverages the split learning paradigm and variational autoencoders to generate high-quality scenarios without moving stakeholder data. With the help of largescale, distributed memory-based experiments, we demonstrate the broad applicability of SplitVAEs in three distinct domain areas: power systems, industrial carbon emissions, and supply chains. Our experiments indicate that SplitVAEs can learn spatial and temporal interdependencies in large-scale networks to generate scenarios that match the joint historical distribution of stakeholder data in a decentralized manner. Our results show that SplitVAEs outperform conventional state-of-the-art methodologies and provide a superior, computationally efficient, and privacy-compliant alternative to scenario generation.

WC21

Summit - 341

Health Care Resource Planning

Contributed Session

1 - Identifying and Understanding the Socioeconomic Factors that have an impact on Hospital Readmissions Rates

Aindrila Chakraborty, Metro State University, St Paul, MN, United States, Jaymeen Shah

Hospital readmissions is a major issue in the healthcare industry today, serving as a crucial indicator of the quality of care patients receive and the related costs. Hospitals are subject to significant penalties for exceeding the expected levels of readmissions each year. Furthermore, studies have shown that a considerable proportion of these readmissions are preventable. Therefore, identifying and understanding the factors leading to hospital readmissions and predicting future occurrences based on these factors are essential tasks for hospitals aiming to lower their readmission rates and enhance patient care quality. Significant research has been conducted to evaluate the effectiveness of programs designed to reduce hospital readmissions by penalizing hospitals for excess rates. Several other research has focused on chronic diseases as a major cause of readmissions. However, there remains a lack of comprehensive studies examining how socioeconomic status and insurance factors contribute to the readmission rates. Our research employs a data-driven approach to explore these socioeconomic drivers using a nationwide dataset. Analyzing the readmission issue from a data analytics perspective is critical as this is a complex interdisciplinary issue as shown in the literature that requires a combined efforts of data analysts besides physicians, healthcare providers, and insurance companies. The findings from our study are intended to provide valuable insights for public health authorities and policymakers, guiding the development of policies and strategies for better treatment and post-discharge care that could decrease readmission rates and associated healthcare costs.

2 - Infection Aware Nurse Staffing

Buyun Li, Kelley School of Business; Indiana University, Bloomington, IN, United States, Jonathan Helm, Pengyi Shi, Kurt Bretthauer

Problem Definition: During outbreaks of infectious diseases, hospitals encounter significant challenges in making well-informed decisions regarding nurse staffing. The crisis in nurse staffing during these outbreaks arises from a simultaneous increase in patient admissions due to the outbreaks and a notably elevated rate of nurse absenteeism caused by infections. The unobservable nature of the infection time and incubation period for these infectious diseases adds complexity to understanding when and how nurses are infected. Lack of this critical information restricts hospital managers from deploying effective and informed operational strategies and staffing plans, thereby limiting their ability to proactively address the staffing crisis.

Methodology/Results: Utilizing data from the IU-Health hospital system during the COVID-19 pandemic, we first develop a Cox proportional hazard model to identify important clinical and operational features for predicting nurse absenteeism caused by infections. Then, we extend standard random graph models to handle the hidden status of nurse infections to capture the disease transmission dynamics with specific probabilistic inference on unobservable/unknown time points of when a nurse is infected or starting to show symptoms. Our model identifies infections from different sources and locations. We gain insights from our model by evaluating mitigation and staffing policies through counterfactual analysis.

Managerial Implications: We find that hospitals can effectively lower the number of nurse infections by designating dedicated units for infectious patients, providing regular testing for infections, and promoting vaccinations. Additionally, staffing decisions play an important role in nurse infection prevention. An ideal staffing level could lower infections by almost 30%.

3 - Predicting All-Types Hospital-Acquired Conditions for Working Age Adults with Multimorbidity Patterns Accounting for Individual Disparity and Structural Racism Effects

Ajit Appari, Boston University, Boston, MA, United States

Hospital-acquired conditions (HACs) are common, costly, and national patient safety priority. Patients with multimorbidity (presence of two or more chronic conditions) require complex care during hospitalization posing significant challenges for clinicians. Prior research often focused on building machine learning models for specific or a subset of HACs among adult population with specific anchor chronic disease and/or adjusting for Charlson or Elixhauser comorbidity indexes.

In my study, I focused on building machine learning models for predicting composite risk of multiple hospital acquired conditions (MHAC) among working age adults having different multimorbidity patterns adjusting for individual disparity factors {age, sex, race/ethnicity, insurance}, structural racism factors {Neighborhood Disadvantage (e.g., poverty, unemployment, households on state income support, and higher proportion of Blacks), Neighborhood Affluence (highly educated, high income population) and Hispanic & Immigrant Concentration}, and provider factors (hospital characteristics).

I used all payer administrative data on 1,645,126 hospitalizations (age: 18-64years) during 2015Q4-2016Q4 across 642 hospitals in Texas. Patients were coded for 69 chronic condition categories (reflecting 4400+ ICD10 diagnosis-and-procedure codes) and grouped into 10 different multimorbidity patterns using unsupervised learning method- Block Clustering with Bernoulli Mixture Models that leverages block classification expectation maximization algorithm. Subsequently, patients were coded for multimorbidity pattern complexity (number of chronic conditions from each patterns) to quantify potential complexity of care needs for presence of such patterns. Finally, I build and evaluate composite risk prediction model for MHACs using Zero-Inflated Negative Binomial Regression, and XGBoost Hurdle Model accounting for multimorbidity pattern complexity, individual disparity factors, structural racism factors, and provider factors.

4 - An Extended Latent Class Choice Model for Characterizing Context-driven hospital choice

Jungwoo Kim, KAIST, Daejeon, Korea, Republic of, Taesik Lee

In the domain of healthcare systems, understanding individuals' choice-making processes is crucial for effective system design and optimization. This is especially pronounced in ambulatory care, where patients are allowed to freely choose a hospital based on their preferences. Hospital choice can be illustrated as context-driven, where explanatory variables affect choices differently depending on the context (e.g., distance's effect varies with sickness severity). To address this, we propose a novel structure of Discrete Choice Model, aiming to provide a comprehensive characterization of context-driven preferences. Our method builds upon the Latent Class Choice Model (LCCM), to better capture the diverse preferences(utility functions) among individuals, improving both interpretability and flexibility. We integrate a membership classification model into LCCM, linking contexts to latent classes of utility functions. Moreover, we propose a structure allowing distinct membership classification models with varying numbers of classes for each subject variable. This approach accommodates different specifications of heterogeneity, enriching details of context-driven preference while also addressing overfitting and data scarcity issues. To estimate this complex structure, we develop an estimation method based on the Expectation Maximization algorithm, with numerical experiments confirming the robustness of our methodology. Our extended LCCM offers comprehensive insights and reliable forecasts for systems featuring complex choice-makers. We anticipate that this capability will provide foundational knowledge when devising policies, interventions, and system modifications to tackle issues in ambulatory healthcare systems and other fields.

WC22

Summit - 342

Energy Consumption and Decarbonization

Contributed Session

Chair: John Paul Pieper

1 - China's Coal Resource Supply-Demand Pattern and "Dual Carbon" Development Strategy Simulation

Li Feng Shi, Tianjin University, Tianjin, China, People's Republic of, Xinyan Jiang, Jiatong Ji, Meixuan Xin, Junying Liu

The energy endowment characteristics of "rich coal, poor oil, and less gas" in China determine the significant role played by the reform of coal resources on the road to achieving the "Dual Carbon" goal. This paper focuses on coal resources, analyzing the supply and demand pattern of China's coal energy products and simulating the development strategy of "Dual Carbon" to aid in achieving the "Dual Carbon" goal. Firstly, starting from the spatial heterogeneity distribution of coal resources, this paper studies the spatial and temporal evolution process of producers and consumers identities in different provinces in the field of coal resources. Secondly, this paper employs the SBM model with undesired output to measure the total factor energy efficiency of each province during the sample period, conducting a comparative analysis based on temporal evolution and regional differences. Finally, this paper selects Shanxi, Shaanxi, and Inner Mongolia, three major coal resource supply provinces of China, as the coal producers, and the Yangtze River Delta, Pearl River Delta, and Beijing-Tianjin-Hebei, three major urban agglomerations of China, as the coal consumers. Building upon this, a system dynamics model of the carbon emission system is constructed. By adjusting key variables in the system, this paper simulates the supply and demand bilateral optimization policy and investigates the impact of policy implementation on China and the Shanxi-Shaanxi-Inner Mongolia region in terms of economy and environment.

2 - Infrastructure planning for cross-border renewable electricity imports: A Stochastic Programming Approach

S Viswanathan, Nanyang Technological University, Singapore, Singapore, Haifeng Qiu , Tobias Massier, Thomas Hamacher

Cross-border renewable electricity imports based on multilateral cooperation can assist countries scarce in renewable energy resources in their roadmap to net zero carbon emissions. In this paper, we develop a two-stage stochastic programming model to determine the optimal infrastructure planning as well as scheduling of importing cross-border renewables with intrinsic short-term intermittency and long-term uncertainty, thus ensuring stable and reliable connection and synchronization of renewable importing sources with the receiving-end power grid from the perspective of technical requirements. Techno-economic evaluation of importing electricity into Singapore from several Southeast Asian countries using the five-years meteorological and load data is carried out in this research. The results indicate that importing electricity through dedicated long-distance cables might be a viable option once the country-level agreements are ironed out. We also perform sensitivity analysis on several of the strict criteria for renewable importation. A new hybrid system planning framework combining both on-site and sharing energy storages demonstrates that leveraging the complementary roles of multi-location importing renewables with a balanced capacity blend of the wind and solar generation breaks through their independent bottlenecks and improves the overall economies. Our results provide a design methodology for power system infrastructure planning for renewable electricity imports, and this could help spawn more renewable energy projects in exporting and importing countries in a win-win fashion.

3 - The Impacts of the Covid-19 Pandemic on the Macroeconomic Efficiency of the OECD Countries

Umit Saglam, East Tennessee State University, Johnson City, TN, United States

In this study, a two-stage Data Enveloped Analysis (DEA) is developed to quantitatively evaluate the impact of the Covid-19 pandemic on macroeconomic efficiencies of the OECD countries by using pre-determined inputs (population, labor, capital stock, and energy consumption) and outputs (real GDP and carbon dioxide emission) variables. The Slack-Based Model (SBM) will be applied to determine undesirable outputs and calculate each country's capital, labor, energy, and emission efficiency. Second, the Malmquist Productivity Index (MPI) is developed to compare productivity changes during the pre-pandemic and pandemic. Third, the sensitivity analysis is conducted to test the robustness of the DEA models. Finally, Tobit regression models are conducted using the second stage analysis's DEA results to investigate the effects of the pandemic. We hope this study's results shed some light on the efficiency assessments of the OECD countries and their productivity changes during the pre-pandemic and pandemic.

4 - Under the goal of carbon neutrality, regional cooperation in reducing emissions can drive economic growth? An empirical examination from the two provinces of Jiangsu and Anhui in China

Jianmin Wang, School of Economics & Management Anhui University of Science and Technology, Huainan, China, People's Republic of, Yang Li

The study proposes emerging issues in the field of carbon neutrality research, that is, whether regional cooperation in emissions reduction under carbon neutrality goals can drive economic growth. Based on the accurate description of the regional economic structure using the 2017 regional input-output table and computable general equilibrium modeling method of Jiangsu and Anhui provinces in China, an empirical investigation was conducted on the economic effects of regional cooperation in emissions reduction under carbon neutrality goals in these two regions. The study found that: (1) With the continuous advancement of regional cooperation in emissions reduction under carbon neutrality goals, multiple economic indicators have been affected and show a systematic expanding trend. (2) Jiangsu and Anhui have different capabilities in dealing with the reduction in fossil fuel use based on their existing economic structures. (3) Regional cooperation in emissions reduction under carbon neutrality goals is advantageous in reducing the economic impacts of dispersed emissions reduction and the absolute economic costs. The study concludes that: (1) Phased development plans for clean energy should be formulated to gradually reduce regional reliance on fossil fuels. (2) Carbon neutrality goals should be steadily promoted based on regional emissions reduction capacity. (3) Accelerating the construction of regional cooperation mechanisms for emissions reduction and establishing a cooperative emissions reduction mechanism under carbon neutrality goals. The research findings may provide experiential insights and policy reference for other countries or regions in the process of achieving carbon neutrality goals while promoting economic growth.

5 - Transportation Decarbonization that is Robust to Global Material Supply Concerns

John Paul Pieper, Carnegie Mellon University, Pittsburgh, PA, United States

Transportation is now the largest source of greenhouse gas (GHG) emissions as well as a major source of local air pollution in the United States, and electrifying vehicles is crucial to decarbonizing the sector. Electric vehicle (EV) batteries contain critical materials that face heterogenous supply chain vulnerabilities due to geographic constraints within up- and mid-stream production. There exists a need for responsive manufacturing to increase firm flexibility in short run EV production to minimize production costs increases due to supply disruptions of critical materials.

We study the value of firm flexibility using a case study of EV production that includes multiple options for flexibility to mitigate supply chain disruptions, including switching battery chemistries or stockpiling critical materials within "medium" run production. Our analysis builds on engineering real option and electric vehicle production model literature to provide novel perspectives of flexible options for EV firms. Using real options analysis, this work will quantify the value of flexibility under supply chain uncertainty by modeling the manufacturing costs, the change in expected consumer adoption of EVs dependent on the performance characteristics, and the risk of undesirable impacts to EV production, manufacturing workers, and U.S. households under different disruption scenarios.

Our preliminary results suggest that stockpiling strategies maintain consumer adoption rates in the presence of disruptions by giving firms options that decrease production costs. We predict battery switching strategies will both maintain EV consumer adoption rates and allow for opportunistic firm behavior by creating more flexible battery production processes.

WC23

Summit - 343

Optimizing Supply Chain and Production Processes

Contributed Session

Contributed

Chair: Shi Qiu, University of Illinois Urbana-Champaign, Urbana, IL, United States

1 - Forecasting Supply Chain Disruption with Heterogeneous Time Series

chaoye pan, Ford Motor Company, Dearborn, MI, United States, Bach Viet Do, Xingyu Li

Operational disruptions have an adverse material impact on the company's performance. Ford operates 37 plants globally, utilizing 17 billion parts annually to manufacture six million cars and trucks. The company has up to 10 tiers of suppliers between itself and raw material. An extended disruption anywhere within this extensive supply chain can inflict substantial financial losses on Ford. The capacity to forecast and identify such disruptions early is crucial to maintain a seamless operation. In the modern era of Big Data, data on suppliers' past performance has increasingly become a valuable source for the analysis and prediction of disruption risks. In our work, we leverage multivariate time series data to forecast supply chain disruptions. This dataset presents significant challenges due to its complexity and vast scale, consisting of over five hundred thousand individual series. The series may exhibit similarities to a certain degree, yet they also demonstrate heterogeneity within specific subgroups. Consequently, traditional industrial and statistical models are limited in capturing the complex dynamics and managing the extensive volume of data. In this presentation, we propose a novel methodology that integrates a modified Attention Sequence to Sequence model Deep Learning Neural Architecture with a Survival Analysis model, to learn intricate data patterns pertaining operational disruptions. Our model has demonstrated a performance with 85% precision and 80% recall during the Quality Assurance (QA) phase across Ford's five North American plants.

2 - Supply Chain Risk Identification: A Practice-based View

Aaron Heinrich, Texas A&M University/Department of Information and Operations Management, College Station, TX, United States, Trevor Hale

Institutions with internalized supply chains regard the finished goods of their operations processes as raw materials they need to produce and deliver the final services and goods they provide customers. For example, a city (an institution) can enforce its laws (a service) for its citizens (its customers) only by sourcing a sustainable supply (an internalized supply chain) of credentialed police officers (both finished goods and raw material inventories) from the police academy it operates (operations process). To identify potential sources of disruption to such supply chains, we analyze empirical data collected from a U.S. nuclear utility that operates an internalized supply chain of credentialed professionals and find its disruptions to be associated with idiosyncrasies in its training programs rather than in its trainees, which upends conventional wisdom. Upon synthesizing two theories from the domains of knowledge management and educational psychology with empirical data collected across multiple industries, we hypothesize and test whether the extent to which training program instructors cognitively apprentice their trainees is related to the extent to which institutions are vulnerable to supply chain disruptions. Our findings permit us to contribute a novel, practice-based view to supply chain risk management and provide actionable insights to operations managers.

3 - Designing An Apparel Closed-Loop Supply Chain Network Using a Robust Multi-Objective Programming

Saman Hassanzadeh Amin, Toronto Metropolitan University, Toronto, ON, Canada, Samira Rouhani , Leslie Wardley

In this talk, designing and optimizing an apparel closed-loop supply chain network is discussed. Several elements such as types of returns, including commercial, end of use, and end of life returns are considered in this network. To design the network, a multi-objective mixed-integer linear programming model based on robust optimization is developed. Then, the results are discussed based on an application in Canada, and the managerial insights are provided. The results show that the proposed model can handle uncertainty under different sustainability objective functions.

4 - The Supply Chain Pressure Index (SCPI): A Predictive Analytics Tool for Agile Supply Chain Risk Management

sein kim, Kyungpook National University, DAEGU, Korea, Republic of, Sungsu Kim

The Supply Chain Pressure Index (SCPI) is an innovative analytical tool developed to measure and mitigate pressures within global supply chains for supply chain risk management, focusing particularly on potential stress factors such as supply-demand imbalances, logistical bottlenecks, and market volatility. This focus sharply contrasts with the broader scope of the Global Supply Chain Pressure Index (GSCPI), which primarily assesses traditional pressure metrics. The proposed SCPI enables more adaptive strategies to changes than the GSCPI's long-term pressure assessments. Additionally, the SCPI integrates various new data sources, including social media, sensor data, and market trends, capturing previously overlooked pressure elements within the supply chain. By incorporating advanced analytics, the SCPI would provide targeted insights that accurately identify critical vulnerabilities and guide immediate corrective actions, significantly reducing risks and enhancing quicker responses to supply chain pressures. The goal of this study is to establish the SCPI as the global standard for agile supply chain risk assessment. Its adoption is expected to lead to significant advancements in global supply chains by integrating advanced diagnostics to proactively manage and mitigate supply chain pressures.

Keywords: Supply Chain Pressure Index (SCPI), Supply Chain Risk Management, Advanced Analytics, Agile Supply Chain, Risk Assessment

5 - Leveraging Blockchain Technology to Mitigate the Bullwhip Effect in Supply Chains: A Theoretical Model and Empirical Analysis

Shi Qiu, University of Illinois Urbana-Champaign, Urbana, IL, United States, Sridhar Seshadri

The bullwhip effect, which manifests as increasing demand variability up the supply chain, creates significant inefficiencies in supply chain operations. This paper investigates the potential of blockchain technology to mitigate this phenomenon. We develop a theoretical model that incorporates demand uncertainty and imperfect demand prediction, examining how blockchain's attributes—such as enhanced transparency, immutability, and smart contracts—can improve information sharing and coordination among supply chain participants. The model posits that blockchain can reduce demand amplification by providing real-time, accurate data and fostering trust among parties. We empirically test our model using observational data from supply chains that have implemented blockchain solutions. The results demonstrate that blockchain adoption can significantly decrease demand variability and enhance supply chain performance. This study contributes to the Operations Management literature by offering a novel perspective on addressing the bullwhip effect through blockchain technology.

WC24

Summit - 344

Optimizing Delivery and Logistics with AI and Crowdsourcing

Contributed Session

Chair: Yooneun Lee, University of Dayton, 300 College Park, Dayton, OH, 45469, United States

1 - A Multi-agent Reinforcement Learning-based Order Dispatching in Delivery Platforms with Diverse Strategic Couriers

Jaewon Choi, Pohang University of Science and Technology, Pohang, Korea, Republic of, Daeho Kim, Dong Gu Choi

Efficiently dispatching orders to couriers has become one of the core technologies of modern on-demand delivery platforms. Particularly, the problem of allocating dynamically occurring requests to multiple service providers has been covered in the field of management science and operations research. However, the diversity of service providers and their strategic decision-making process has not been fully considered. In a realistic delivery platform environment, couriers have the rights to reject assigned orders, and the platform operator must take this into account while dispatching orders. Also, couriers feature different performance capabilities depending on their proficiency and this heterogeneity influences their acceptance behavior. We develop a framework to intelligently dispatch orders within a delivery platform based on a multi-agent Markov decision process model considering strategic couriers. We then verify the effectiveness of our logic in a dynamic simulation environment based on real-life data.

2 - End-to-End Trip to Driver Dispatchment Timing of Scheduled Delivery: A General Decision-Making System for Crowdsourced Delivery Platforms

Soumita Saha, Walmart Global Tech India, Bangalore, India, Tao Cao, Ankit Dhokariya, Weihang Ren, Minghui Liu, Yuan Wang, Sherry Wan, Jing Huang, Mingang Fu, Prakhar Mehrotra, Chittaranjan Tripathy

Crowdsourcing delivery platforms used to face most demand from real-time on-demand requests. Recently, with increasing adoption by ecommerce retailers, we observed a shift of demand pattern from on-demand to scheduled delivery requests. The increasing reliance on crowdsourced drivers for scheduled deliveries presents a new challenge. That is, the timings through the process of dispatching trips to drivers set the foundation of the delivery journey. Despite the critical role of end-to-end dispatch timing decisions, there has been a paucity of holistic approaches to this decision-making process that ensure cost-efficient, on-time deliveries. We present a novel system designed to optimize the end-to-end driver dispatch timing decisions. The system begins by predicting time durations of segments throughout delivery journeys. It then utilizes simulation, stochastic programming, and survival regression modeling to optimally determine arrival time of drivers, start time of driver dispatch, and initialization time of surge pay. The objectives are maximizing time and cost efficiency in dispatch process while meeting promised delivery time for customers. Our novel decision-making system is adaptable to dynamic market conditions through configurable parameters, which are learned automatically via market segmentations. We have implemented this system on Walmart's proprietary crowdsourced last-mile delivery platform. Our simulation and experimentation studies indicate significant improvements in dispatch efficiency, by reducing 55% dispatch idle time compared to previous solutions while maintaining high on-time delivery rates. Our work demonstrates that the framework underlying the system has the potential to enhance delivery experiences across other crowdsourced delivery platforms, where on-time delivery promise for customers are foundational.

3 - The Multichannel Last-Mile Delivery Problem with Crowdsourcing Integration

Sang Jin Kweon, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea, Republic of, Jaesung Kim

In recent years, crowdsourcing has gained attention in last-mile logistics. This study addresses the multichannel last-mile delivery Problem integrating crowdsourcing options. We propose a delivery pricing strategy to incorporate individual crowdsourced fleets into the platform. Our mathematical model optimizes terminal locations for delivery-dedicated and individual crowdsourced fleets, aiming to minimize total last-mile delivery costs amid fluctuating demand and pricing based on a city's floating population. Experimental findings demonstrate significant cost reductions by strategically deploying crowd workers in densely populated areas rather than merely proximity to final delivery destinations. Additionally, we examine the cost and carbon emissions implications of integrating crowdsourcing and deploying electric vehicles for last-mile deliveries.

4 - Optimizing Order Consolidation and Vehicle Routing for Multi-Warehouse Retailers: A Mixed Integer Linear Programming Approach

Yooneun Lee, University of Dayton, Dayton, OH, United States, Nikesh Chithambaram, Seokgi Lee, Hyeong-Suk Na

Order consolidation has emerged as a viable solution to address the last-mile delivery challenges faced by online retailers dealing with higher volumes and frequencies of customer orders. However, making order consolidation decisions is complex due to various order time frames, the diverse range of items managed and fulfilled by multiple warehouses, and the absence of efficient algorithms to synchronize consolidation decisions with vehicle routing and dispatching. This study presents order consolidation policies and addresses the vehicle routing problem for a retailer operating multiple warehouses, aiming to simultaneously optimize order fulfillment and shipment costs. The problem is formulated as a mixed integer linear programming (MILP) model. The proposed scenarios for solving this problem are analyzed to provide practical recommendations. Furthermore, the study compares the results with those obtained from a reinforcement learning algorithm developed in a previous study, demonstrating the efficiency of the heuristic algorithm.

WC25

Summit - 345

Economic Strategies for a Sustainable Future

Contributed Session

Chair: Takuro Murakami, Keio University, Yokohama, N/A, Japan

1 - Forecasting the Success of International Joint Ventures

Aws Hamo Younes, Durham University , Durham, United Kingdom, Konstantinos NIKOLOPOULOS, Michel Phan

A joint venture is a business partnership where two or more companies work together on a specific project, sharing resources and risks. Each company keeps its own identity while collaborating to achieve a common goal. Partner selection can be defined as the process of seeking, evaluating, and finally choosing the right partner to achieve the firm's strategic growth objectives in a specific host country. Country Governance is defined as the traditions and institutions by which authority in a country is exercised. There are six dimensions of governance are constructed based on this definition; Voice and Accountability (VA), Political Stability (PV), Government Effectiveness (GE), Regulatory Quality (RQ), Rule of Law (RL) and Control of Corruption (CC).

In an increasingly globalized world, firms aiming to stay competitive in international business must explore foreign markets and many multinational enterprises (MNEs) from developed countries use Foreign Direct Investment (FDI) as a strategy to enter markets in developing economies, benefiting from local knowledge and connections. Due to external uncertainties and the challenges of operating in a foreign market, MNEs often prefer forming International Joint

Ventures (IJVs) with local partners rather than wholly-owned subsidiaries (WOSs) to enhance the success of their international endeavours and reduce the risk of failure. This study focuses on analysing how weak country governance influences the criteria MNEs use to select local partners and develops a forecasting model for Forecasting the success of International Joint Ventures.

2 - Assessing Combined Innovation Potential of Firms through Value-Guided Transformers: Implications for Post-M&A Outcomes

Naveenkumar Ramaraju, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Gautam Pant

We propose a framework to assess the combined innovation potential of firms and validate it by evaluating its impact on the post-M&A outcomes of such firms. We leverage the "Value Innovation Theory" for the framework design, which suggests that innovation should be value-driven rather than technology-driven. Specifically, we propose value embeddings (i.e., value-based vector representations) to encode the value of firm patents in the semantic-value space. An information artifact called CoMET (Combined Maximum Embedding Total) is constructed from value embeddings of patents to measure the combined innovation potential of firm pairs. We evaluate the proposed information artifact in the context of mergers and acquisitions (M&A). We find that M&As with higher combined innovation potential of constituent firms, as measured through CoMET before mergers, have higher innovation quantity, quality, and knowledge-sharing post-merger. Heterogeneity analyses show that our framework to measure the combined innovation potential is a useful indicator of post-M&A innovation output beyond the traditional indicators such as interfirm technology similarity and complementarity.

3 - Profit sharing between differentiated firms in open innovation system

Takuro Murakami, Keio University, Yokohama, Japan, Nobuo Matsubayashi

We consider an open innovation system in which multiple third-party firms are invited to participate in a platform to create an innovative product. Our focus is on the platform's profit-sharing scheme with these participants, who may differ from each other asymmetrically in terms of complementarity and substitutability. Specifically, we compare three schemes: equal splitting, profit sharing proportional to effort level, and profit sharing proportional to effort level adjusted by the degree of complementarity and substitutability. We develop a game-theoretic model where a monopolistic platform firm offers one of these profit-sharing schemes, and the third-party participants determine their effort levels to maximize their individual profits. Our equilibrium analysis reveals that the optimal profit-sharing scheme depends on the complementarity-substitutability relationship of participants, and the cost structure of innovation.

WC26

Summit - 346

Causal Inference and Machine Learning in OM

Invited Session

MSOM: Service Operations

Chair: Ruomeng Cui, Emory University, Amazon, Decatur, GA, United States

Co-Chair: Zhikun Lu, Emory University, Decatur, GA, United States

1 - The Long-Term Impact of Wage on Workers&Rsquo; Behaviors in Sharing Economy

Shilei Luo, Washington University in St. Louis, St. Louis, MO, United States, Xinyu Cao, Dennis Zhang

In this paper, we examine the long-term effects of hourly wages on gig workers' behavior within the sharing economy through a field experiment that randomizes gig workers' hourly pay. We analyze the dynamics of these wage impacts on the workers' decision to join the platform and the quantity of the services they supply. Our findings reveal that higher wages lead to an increase in the quantity of service supplied and conducted, with the impact on service conducted intensifying over time. This dynamic is largely attributable to the workers' reputation growth and the accumulation of repeat customers.

2 - Policy Learning Under Biased Sample Selection

Roshni Sahoo, Stanford University, Stanford, CA, United States, Lihua Lei, Stefan Wager

Practitioners often use data from a randomized controlled trial to learn a treatment assignment policy that can be deployed on a target population. A recurring concern in doing so is that, even if the randomized trial was well-executed (i.e., internal validity holds), the study participants may not represent a random sample of the target population (i.e., external validity fails)--and this may lead to policies that perform suboptimally on the target population. We consider a model where observable attributes can impact sample selection probabilities arbitrarily but the effect of unobservable attributes is bounded by a constant, and we aim to learn policies with the best possible performance guarantees that hold under any sampling bias of this type. In particular, we derive the partial identification result for the worst-case welfare in the presence of sampling bias and show that the optimal max-min, max-min gain, and minimax regret policies depend on both the conditional average treatment effect (CATE) and the conditional value-at-risk (CVaR) of potential outcomes given covariates. To avoid finite-sample inefficiencies of plug-in estimates, we further provide an end-to-end procedure for learning the optimal max-min and max-min gain policies that does not require the separate estimation of nuisance parameters.

3 - The Value of Last-Mile Delivery in Online Retail

Ruomeng Cui, Emory University, Amazon, Decatur, GA, United States, Zhikun Lu, Tianshu Sun, Lixia Wu

Last-mile delivery refers to the final leg of shipment in which packages are moved from a local transportation hub to end-customers. It is known to be the most expensive part of delivery, due to operations complexity and lack of economy of scale. Given the growing importance of last-mile delivery in online retail, companies need to decide whether to outsource this service to customers through pickup stations where they can collect packages or to offer home delivery. In this paper, we study the economic value of last-mile delivery. We conducted a quasi-experiment in collaboration with Cainiao, Alibaba's logistics subsidiary, in which home delivery service was rolled out to some pickup stations sequentially in 2021. This allowed us to comprehensively evaluate the causal impact of last-mile delivery. Using a staggered difference-in-differences identification, we found that last-mile delivery significantly increases sales and customer spending on Alibaba's retail platform. Last-mile home delivery relies heavily on physical labor and thus has capacity constraints. To optimally prioritize that capacity, we used an uplift model with causal machine learning to target the most responsive customers. Specifically, we proposed a novel capacity- and fairness-aware uplift model that builds in capacity and fairness constraints to optimize the targeting policy, thereby maximizing profits without compromising social equity.

4 - Dynamic Local Average Treatment Effects

Ravi Sojitra, Stanford University, Stanford, CA, United States, Vasilis Syrgkanis

We consider Dynamic Treatment Regimes (DTRs) with one sided non-compliance that arise in applications such as digital recommendations and adaptive medical trials. These are settings where decision makers encourage individuals to take treatments over time, but adapt encouragements based on previous encouragements, treatments, states, and outcomes. Importantly, individuals may choose to (not) comply with a treatment recommendation, whenever it is made available to them, based on unobserved confounding factors. We provide non-parametric identification, estimation, and inference for Dynamic Local Average Treatment Effects, which are expected values of multi-period treatment contrasts among appropriately defined complier subpopulations. Under standard assumptions in the Instrumental Variable and DTR literature, we show that one can identify local average effects of contrasts that correspond to offering treatment at any single time step. Under an additional cross-period effect-compliance independence assumption, which is satisfied in Staggered Adoption settings and a generalization of them, which we define as Staggered Compliance settings, we identify local average treatment effects of treating in multiple time periods.

WC27

Summit - 347

Yichun Hu session

Invited Session

Applied Probability Society

Chair: YICHUN HU, Cornell University, Ithaca, NY, 14850, United States

1 - Joint Probability Estimation of Many Binary Outcomes via Localized Lasso

Yan Chen, Duke University, Durham, NC, United States, Alexandre Belloni , Matthew Harding

In this work we consider estimating the probability of many (possibly dependent) binary outcomes which is at the core of many applications, e.g., multi-treatment in causal inference, demands for bundle of products, etc. Without further conditions, the probability distribution of an M dimensional binary vector is characterized by exponentially in M coefficients which can lead to a high-dimensional problem even without the presence of covariates. Understanding the (in)dependence structure allows to substantially improve the estimation as it allows for an effective factorization of the probability distribution. In order to estimate the probability distribution of a M dimensional binary vector, We leverage a Bahadur representation that connects the sparsity of its coefficients with independence across the components. We propose to use regularized estimators to obtain an adaptive estimator with respect to the dependence structure which allow for rates of convergence to depend on this intrinsic (lower) dimension. Our main results consider the presence of (low dimensional) covariates for which we propose a locally penalized estimator. We provide point-wise and uniform rates of convergence addressing several issues in the theoretical analyzes as we strive for making a computationally tractable formulation. We apply our results in the estimation of causal effects with multiple binary treatments and show how our estimators can improve the finite sample performance when compared with non-adaptive estimators that try to estimate all the probabilities directly. We also provide simulations that are consistent with our theoretical findings.

2 - Survey Descent: a Case-Study in Amplifying OR Research with Modern ML Workflows

X.Y. Han, Chicago Booth, Chicago, IL, United States

Within the classic operations research (OR) curriculum, one learns that for strongly convex objectives that are smooth, gradient descent ensures linear convergence of iterates and objective values relative to the number of gradient evaluations. Nonsmooth objective functions are more challenging: existing solutions in the OR literature typically invoke cutting plane methods whose complexities are difficult to bound, leading to convergence guarantees that are sublinear in the cumulative number of gradient evaluations. We instead propose a multipoint generalization of the gradient descent called Survey Descent. In this method, one first leverages a one-time initialization procedure to gather a "survey" of points. Then, during each iteration of the method, the survey points are updated in parallel using a simple, four-line procedure inspired by gradient descent. Under certain regularity conditions, we prove that Survey Descent then achieves a desirable performance by converging linearly to the optimal solution in the nonsmooth setting.

Despite being an OR endeavor, we discuss how the development of Survey Descent was significantly accelerated by a frictionless computational workflow made possible by tools from modern machine learning (ML); how this model of applying new ML workflows to solve open questions in optimization and applied probability could amplify the productivity of OR research; practical computational bottlenecks that could hinder this integration; and what tools are needed to overcome those obstacles.

3 - RL Algorithm Design for Improving Post-HCT Medication Adherence via a Digital Intervention

Ziping Xu, Harvard University, Allston, MA, United States, Susan Murphy

The survival of adolescents and young adults (AYA) who undergone hematopoietic stem cell transplantation critically relies on the adherence to medication that mitigates acute graft-versus host diseases. For AYAs with cancer, self-management rarely involves the individual alone and their care-partner often plays an important role, which forms a dyadic structure. The oversight of this dyadic structure may partially explain small and heterogeneous effects of previous interventions aimed at AYA adherence. In this paper, we design RL algorithms that provide personalized and just-in-time digital interventions on AYA, care-partner and their relationship to improve AYA's medication adherence. We construct a simulation testbed from an offline dataset combined with domain experts' knowledge. The testbed is used to select a set of candidate algorithms designed for the dyadic structure. Our results indicate that the appropriate choice of RL algorithm design including pooling and algorithm model can significantly improve the cumulative rewards.

4 - Property Test on the Optimal Assortment in the Contextual Multinomial Logit Model with Adaptive Sampling

Shuting Shen, Duke University, Durham, NC, United States, Alexandre Belloni, Ethan Fang

We study the combinatorial inference of the optimal assortment within the framework of the contextual multinomial logit model, where we assume that the observed data are collected over time, with contextual information varying temporally. Moreover, the offer set provided at each time point is adaptively estimated based on history data. We propose an inferential procedure that constructs a discrete confidence set for the latent optimal assortment, enabling us to infer properties of the optimal assortment, such as the number of product categories to include in the optimal offer set. The temporal dependency in the discrete supports of the summands over the products for the Hessian matrix of the log-likelihood function brings challenges to showing its convergence. To address this, we have developed novel anti-concentration bounds for the difference of the maxima of two Gaussian random vectors. Additionally, we tackle the high dimensionality of the combinatorial inference problem by utilizing subspace projection techniques. We offer theoretical guarantees for both the validity and effectiveness of our inferential procedure and establish the information-theoretic lower bounds for the necessary signal strength.

WC28

Summit - 348

Optimization of Layout and Stocking

Contributed Session

Contributed

Chair: Tejaswini Samal, Indian Institute of Technology, Kharagpur

1 - Biasness in ordering: An experiment with varying holding and backorder costs

Sourish Sarkar, Penn State Erie, Erie, PA, United States

The optimal inventory level depends on the holding and backordering costs. However, past behavioral studies considered these costs fixed, and didn’t investigate the effect of such costs on the underweighting bias. A controlled laboratory experiment allows us to explore the impact of varying these costs on the cognitive bias that contributes to the bullwhip effect.

2 - Strategic Communications in Entry to a Joint Innovation Venture

Zhenxiao Chen, Southern University of Science and Technology, Shenzhen, China, People's Republic of, Kanglin CHEN, Qiaochu He

This paper depicts a frenemy relationship between an incumbent and a potential entrant in developing an innovative product with uncertain market potential. The entrant contributes to the product development but encroaches on the monopoly of the incumbent. Regarding the market entry decision, we find that an entrant with a higher market power may be less inclined to encroach on the market. In cases where the entrant has already encroached on the market, the two parties jointly exert efforts only when the synergy efficiency is high and the two parties are comparable in terms of market power; otherwise, the party with a limited market power becomes a free-rider. To reconcile the conflicting interests between both parties, we explore how the incumbent can communicate the market potential with the entrant strategically under a framework of information design. The optimal information scheme features three different signal structures: a downplay policy that deters entry, a hype policy that facilitates entry, and a silent policy that carries no information. The specific signal structure's choice is contingent upon the value of the new entrant and the prior expectation about the market potential. From the overall market perspective, an entrant comparable in market power optimally benefits the incumbent and the entire market. In that case, the two parties are highly incentivized to develop the product by exerting considerable efforts, which fully exploits the complementary effect and reinforces the investment by the parties.

3 - Relatively Robust Multiproduct Pricing

Ru Zhang, EPFL, Lausanne, Switzerland, Thomas Weber

This study investigates the robust pricing of multiple products when demand information is limited to parameter intervals, accommodating general nonlinear demand systems. The optimal robust price is obtained by maximizing a performance index, representing the worst-case ratio of attained profit to the optimal ex-post profit achievable in the absence of model ambiguity. If the firm’s product portfolio contains at least one profitable item, the optimal performance index offers a meaningful relative performance guarantee. Computational simplification is substantial due to the "envelope property" for linear multiproduct demand systems, allowing the performance index to be expressed as a function of just two vertices in the parameter space. Indeed, the linear multiproduct demand system with interval coefficients is fundamentally compatible with any Lipschitz-continuous nonlinear demand system, as the latter can be linearized around any relevant operating point. Thus, the model can be identified using appropriate OLS confidence intervals for the coefficients of a linear demand system. Regarding optimal robust pricing, in certain symmetric settings with linear, semi-log, or log-log demand models, closed-form solutions are available. For a large sample of random instances of the multiproduct pricing problem we compare our solution, based on minimizing maximum relative regret, to solutions obtained using alternative criteria, such as maximin payoff, minimax absolute regret, and Laplacian uncertainty. The findings demonstrate superior performance of our solutions, validated also using large-scale real-world data from an online retailer.

4 - Deep Reinforcement Learning based Ensemble Approach for Inventory Optimization with multi-item news vendor problem

Tejaswini Samal, Indian Institute of Technology, Kharagpur, Kharagpur, India, Anupam Ghosh

This paper addresses a multi-item newsvendor problem with both capacity and budget constraints. While this problem has been previously studied in the literature using the classical newsvendor problem and its variations, existing solutions often rely on assumptions about demand distributions that can lead to errors in inventory optimization. However, existing solutions often relied on assumptions about demand distributions, which are often incorrect and led to errors in the inventory optimization. To address this issue, this research proposes a distribution-free and entirely data-driven approach. Utilizing sample demand data as input, the proposed approach employs a deep reinforcement learning based ensemble approach with empirical risk minimization principle to determine optimal order quantities. Additionally, a heuristic is developed using hierarchies of the retail products to perform multi-item inventory optimization under both capacity and budget constraint. The proposed approach is validated on a real-world dataset of retail products. Comparative analysis of the results against data driven max-min and empirical inventory optimization methods demonstrates the superior performance of the proposed approach. Furthermore, total inventory cost comparisons reveal a significant reduction attributable to the proposed data-driven methodology.

WC29

Summit - 420

Vehicle Routing Problems & Co.

Invited Session

OPT: Integer and Discrete Optimization

Chair: Nicholas Parham, Air Force, 10476 Glenbrittle Ave, Las Vegas, 89166, United States

1 - Introducing a mixed-integer nonlinear program to the air tasking cycle

Nicholas Parham, Air Force, Las Vegas, NV, United States

The air tasking cycle is an iterative battle rhythm that involves creating objectives, developing targets, weaponeering targets, producing orders, executing orders, and assessing execution. Producing orders, in the form of a schedule, is a key purpose of the tasking cycle. This schedule coordinates capabilities to create desired effects in support of joint force objectives. The scheduling problem that arises can be formulated as a rich vehicle routing problem, which is a variant that incorporates multiple types of constraints to handle real-life considerations. We present a mixed-integer nonlinear program that prescribes where aircraft need to be, at what time, and with which loadout, so as to maximize support of joint force objectives. Demonstrations of this model on synthetic instances show that the total man-hours spent planning can be significantly reduced at no cost to schedule quality.

2 - TBD

Lucas Sippel, The University of Queensland, Brisbane, Australia

TBD

3 - A Two-Echelon Vehicle Routing Problem with Mobile Satellites and Multiple Commodities

Aria Dahimi, Eindhoven University of Technology, Eindhoven, Netherlands, Virginie Lurkin, Mehrdad Mohammadi, Tom Van Woensel

The exponential growth of e-commerce has led to a surge in last-mile delivery demands, presenting intricate challenges for logistics service providers. In this study, we introduced the Two Echelon Vehicle Routing Problem (2E-VRP) with mobile satellites, multiple depots, and multiple commodities (3M-2E-VRP). A central focus of our work is exploring direct delivery using the first echelon vehicles (mobile satellites) and the potential for exchanging goods at meeting points, which can be specific locations such as parking lots or customer locations. To tackle this problem, we develop a mixed-integer linear programming (MILP) model and a math-heuristic algorithm named Approximated Learning based Large Neighbourhood Search (AL-LNS). The AL-LNS algorithm integrates innovative approximated and exact scheduling techniques to navigate within a learning-based procedure, all set up within a Large Neighborhood Search (LNS) structure. Computational experiments evaluate the validity of the 3M-2E-VRP formulation and the effectiveness of the math-heuristic. Finally, a sensitivity analysis was conducted to investigate the impact of input parameters on the final delivery solution, providing precious insights for decision makers and logistic service providers.

4 - An iterative sample scenario approach for the dynamic dispatch waves problem

Sandjai Bhulai, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, Leon Lan, Jasper van Doorn, Niels Wouda, Arpan Rijal

Same-day delivery operations face the challenge of unknown, dynamically revealed delivery requests, necessitating a balance between immediate vehicle dispatch for timely delivery and delayed dispatch to integrate future requests. We explore the dynamic dispatch waves problem where vehicles are dispatched at predetermined times. At these times, operators must decide which requests to dispatch and their routing to minimize total costs while ensuring timely service. We introduce an iterative conditional dispatch (ICD) strategy, using a sample scenario method to categorize requests into dispatch, postpone, or undecided groups, with undecided requests diminishing each iteration until a final decision in the last. ICD comes in two forms: one uses thresholds, the other leverages similarity for decisions. Notably, ICD's simplicity, conceptually straightforward and easy to implement, does not detract from its effectiveness. Extensive numerical experiments show both ICD variants adeptly handle large decision spaces and quickly reach high-quality solutions.

WC30

Summit - 421

Interacting Particle Systems and Stochastic Networks

Invited Session

Applied Probability Society

Chair: Sayan Banerjee, University of North Carolina, Chapel Hill, Cary, NC, United States

1 - Online Heavy-tailed Estimation and Change-point Detection

Abishek Sankararaman, Amazon, Seattle, WA, United States

Motivated by anomaly detection applications in cloud computing, this talk considers two theoretical questions on online change-point detection and mean estimation and shows that clipped-Stochastic Gradient Descent (clipped-SGD) is both a practical algorithm that enjoys good theoretical guarantees. Both of these problems are common sub-routines in variety of cloud computing applications ranging from monitoring, anomaly detection to control. The online change-point detection (OCPD) problem is one in which samples that are potentially heavy-tailed, are presented one at a time and a change in the underlying mean must be detected as early as possible. For this OCPD problem, we show that clipped-SGD enjoys finite sample guarantees on False positives and negatives and an upper bound on the worst case detection delay. This is the first bound for OCPD that works simultaneously across all settings - high-dimensional data that is either heavy-tailed, light-tailed or discrete distribution. The online estimation problem is one where a certain parameter of interest needs to be estimated from a high-dimensional, heavy-tailed and corrupted data-streams. For online estimation tasks that can be cast as minimizing a strongly convex loss function, we prove that an appropriately tuned version of the clipped-SGD is simultaneously {(i)} adaptive to drift, { (ii)} robust to heavy-tailed inliers and arbitrary corruptions, {(iii)} requires no distributional knowledge and {(iv)} can be implemented in an online streaming fashion. All prior estimation algorithms have only been proven to posses a subset of these practical desiderata. The talk is based in part on the following three papers

https://papers.nips.cc/paper_files/paper/2023/file/9e15d892c63903ecc278e0dd05536951-Paper-Conference.pdf

https://arxiv.org/abs/2306.09548

https://proceedings.mlr.press/v162/sankararaman22a/sankararaman22a.pdf

2 - High-dimensional scaling limits and fluctuations of online least-squares SGD with smooth covariance

Krishnakumar Balasubramanian, University of California, Davis, Davis, CA, United States

We derive high-dimensional scaling limits and fluctuations for the online least-squares Stochastic Gradient Descent (SGD) algorithm by taking the properties of the data generating model explicitly into consideration. Our approach treats the SGD iterates as an interacting particle system, where the expected interaction is characterized by the covariance structure of the input. Assuming smoothness conditions on moments of order up to eight orders, and without explicitly assuming Gaussianity, we establish the high-dimensional scaling limits and fluctuations in the form of infinite-dimensional Ordinary Differential Equations (ODEs) or Stochastic Differential Equations (SDEs). Our results reveal a precise three-step phase transition of the iterates; it goes from being ballistic, to diffusive, and finally to purely random behavior, as the noise variance goes from low, to moderate and finally to very-high noise setting. In the low-noise setting, we further characterize the precise fluctuations of the (scaled) iterates as infinite-dimensional SDEs. We also show the existence and uniqueness of solutions to the derived limiting ODEs and SDEs. Our results have several applications, including characterization of the limiting mean-square estimation or prediction errors and their fluctuations, which can be obtained by analytically or numerically solving the limiting equations.

3 - Moderate deviation principles for Join-the-Shortest-Queue-d system

Ruoyu Wu, Iowa State University, Ames, IA, United States, Zhenhua Wang

We consider the large-scale load balancing queueing system under the Join-the-Shortest-Queue-d policy. The system consists of one dispatcher and n servers. When a task arrives at the dispatcher, d servers are chosen uniformly at random, and then the task is routed to the one with the shortest queue length. We establish moderate deviation principles for the system occupancy process and average queue length process, formulated in terms of large deviation principles for the centered and scaled processes with an appropriate speed function. Proofs rely on certain variational representations for exponential functionals of Poisson random measures and weak convergence arguments. This is joint work with Zhenhua Wang.

4 - Reflected Brownian Motion with Drift in a Wedge and the Submartingale Problem

Ziran Liu, New York University, New York, NY, United States, Peter Lakner, Josh Reed

We study reflecting Brownian motion with drift constrained to a wedge in the plane. Our first set of results provides necessary and sufficient conditions for existence and uniqueness of a solution to the corresponding submartingale problem with drift, and show that its solution possesses the Markov and Feller properties. Next, we study a version of the problem with absorption at the vertex of the wedge. In this case, we provide a condition for existence and uniqueness of a solution to the problem and some results on the probability of the vertex being reached.

WC31

Summit - 422

Learning for Contextual Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Ian Zhu, NUS Business School, Singapore, Singapore

1 - Multiproduct Inventory and Pricing with Contextual Robust Optimization

Qinshen Tang, Nanyang Techonological University, Singapore, Singapore, Xun Zhang, Zhi Chen, LI CHEN

Traditionally, multiproduct inventory and pricing problems are approached by initially estimating parameters for a presumed "sufficiently accurate" demand model, followed by optimizing specific models to determine optimal inventory and pricing decisions. However, obtaining an accurate demand model is nearly impossible due to unobservable parameters (parameter uncertainty) and the unknown distribution of the error term in the stochastic demand model (residual ambiguity). Additionally, the predicted demand is endogenously linked with pricing, leading to decision-dependent predictions that often result in intractable bilinear optimization problems. This paper addresses these challenges by introducing a contextual robust optimization (CRO) model to tackle both issues. Despite the intractability of the CRO model, we propose a fortified affine recourse approximation to resolve the decision-dependent prediction issue, reformulating the problem as a semidefinite programming model. Our extensive numerical studies demonstrate the effectiveness of the CRO model, outperforming the conventional predict-then-optimize approach in terms of average profit in out-of-sample tests.

2 - Conformal Inverse Optimization

Timothy Chan, University of Toronto, Toronto, ON, Canada, Bo Lin, Erick Delage

Inverse optimization has been increasingly used to estimate unknown parameters in an optimization model based on decision data. We show that such a point estimation is insufficient in a prescriptive setting where the estimated parameters are used to prescribe new decisions. The prescribed decisions may be low-quality and misaligned with human intuition and thus are unlikely to be adopted. To tackle this challenge, we propose conformal inverse optimization, which seeks to learn an uncertainty set for the unknown parameters and then solve a robust optimization model to prescribe new decisions. Under mild assumptions, we show that the suggested decisions can achieve bounded out-of-sample optimality gaps, as evaluated using both the ground-truth parameters and the decision maker's perception of the unknown parameters. Our method demonstrates strong empirical performance compared to classic inverse optimization.

3 - The Perturbation Gradient Surrogate Loss

Michael Huang, CUNY Baruch College, New York, NY, United States, Vishal Gupta

We propose a novel family of decision-aware surrogate losses, called Perturbation Gradient (PG) losses, for the predict-then-optimize framework. These losses directly approximate the downstream decision loss and can be optimized using off-the-shelf gradient-based methods. Importantly, unlike existing surrogate losses, the approximation error of our PG losses vanishes as the number of samples grows. This implies that optimizing our surrogate loss yields a best-in-class policy asymptotically, even in misspecified settings. This is the first such result in misspecified settings and we provide numerical evidence confirming our PG losses substantively outperform existing proposals when the underlying model is misspecified and the noise is not centrally symmetric. Insofar as misspecification is commonplace in practice -- especially when we might prefer a simpler, more interpretable model -- PG losses offer a novel, theoretically justified, method for computationally tractable decision-aware learning.

4 - Robustness in Inverse Optimization

Honglin An, Institute of Operations Research and Analytics, Singapore, Singapore, Ian Zhu

Inverse optimization has become widely studied and has permeated into a wide variety of applications. As it has been observed in the literature, parameter estimates obtained through inverse optimization can be highly sensitive to model inputs. In this talk, we establish a subtle connection between parameter instability and the robustness of inverse optimization models. In particular, we show that parameter instability can result from a lack of robustness to model inputs, motivating a new robust inverse optimization framework.

WC32

Summit - 423

Advances in Nonlinear Continuous Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Qi Wang, Lehigh University, Bethlehem, PA, United States

1 - Structured Non-Smooth Non-Convex Composite Optimization Without Weak Convexity

YAO YAO, The University of Iowa, Iowa City, IA, United States, Qihang Lin, Tianbao Yang

This paper explores numerical methods for solving a structured non-smooth non-convex composite optimization problem without weak convexity property, which is a composite of a differentiable mapping and a non-convex non-smooth Lipschitz function that allows a simple proximal mapping. With the Moreau Envelope technique, the composite problem can be formulated as a difference-of-convex (DC) program. Then, we introduce single loop deterministic and stochastic algorithms to solve the problem. We show the iteration complexity of the proposed methods for finding an ϵ-Goldstein-stationary point. Moreover, we show that this method can be applied to fairness problem with Wasserstein distance-based fairness metric.

2 - Inexact Sequential Quadratic Programming Method for Nonlinear Constrained Stochastic Optimization

Baoyu Zhou, Arizona State University, Tempe, AZ, United States, Frank E. Curtis, Daniel Robinson

We propose a stochastic sequential quadratic programming algorithm for solving continuous optimization problems with nonlinear constraints. It is assumed that constraint function and derivative values can be computed, but that only stochastic approximations are available for the objective function and its derivatives. Our algorithm only employs stochastic objective gradient estimates without using any objective function values or estimates, and it allows inexact subproblem solutions to be employed, which is particularly useful in large-scale settings. Conditions are imposed on the inexact subproblem solutions that account for the fact that only stochastic objective gradient estimates are employed. Convergence results are established for the method. Numerical experiments show that the proposed method vastly outperforms a stochastic subgradient method and can outperform an alternative SQP algorithm that employs highly accurate subproblem solutions in every iteration.

3 - A Study on the Two-Metric Projection Method

Yue Xie, University of Hong Kong, Pokfulam, Hong Kong, Hong Kong

The two-metric projection method is a simple yet elegant algorithm proposed by Bertsekas in 1984 to address bound/box-constrained optimization problems. The algorithm's low per-iteration cost and potential for using Hessian information makes it a favourable computation method for this problem class. However, its global convergence guarantee is not studied in the nonconvex regime. In our work, we first investigate the global complexity of such a method for finding first-order stationary solution. After properly scaling each step, we equip the algorithm with competitive complexity guarantees. Furthermore, we generalize the two-metric projection method for solving ℓ_1-norm minimization and discuss its properties via theoretical statements and numerical experiments.

WC33

Summit - 424

Applications of Optimization in Statistics

Invited Session

OPT: Optimization Under Uncertainty

Chair: Stanislav Uryasev, Stony Brook University, Stony Brook, NY, United States

1 - EM Algorithms for Optimization Problems with Polynomial Objectives

Jun-ya Gotoh, Chuo university, Bunkyo-ku, Tokyo, Japan, Kensuke Asai

EM (Expectation-Maximization) algorithm is a popular algorithm often used in maximum likelihood estimation of statistical models when they include latent random variables. Focusing on its features as an MM (Majorization-Minimization) optimization algorithm, a new algorithmic scheme based on the EM algorithm is presented for solving general (not necessarily statistical) optimization problems by introducing certain probability distributions. We show that by choosing an appropriate distribution from the exponential family, optimization problems with polynomial objective functions, which include linear program (LP) and quadratic program (QP), can be embedded in the EM scheme. Specifically, with a normal distribution, the EM algorithm for the unconstrained minimization of a convex quadratic function turns out to be a natural gradient descent without line search. For minimizing a polynomial function over a rectangle or a simplex, the EM algorithm with a binomial or multinomial distribution, respectively, becomes an interior point algorithm which is reduced to a natural gradient descent without line search. As further extensions to the case where a natural gradient descent without line search is no longer available, new interior point algorithms are presented for polynomial optimization over a polytope, LP, and QP, based on EM and generalized EM algorithms. The presented algorithms for the extensions are somewhat conceptual in that their sub-procedures require solving a nonlinear convex optimization, and some practical strategies are discussed.

2 - Risk Quadrangle and Robust Optimization Based on phi-divergence

Cheng Peng, Stony Brook University, Stony Brook, NY, United States

Distributionally robust optimization minimizes the maximum expected loss, where the maximum is over an uncertainty set defined by φ-divergence. The maximum expected loss defines the φ-divergence risk measure. This paper studies robust and distributionally robust optimization based on the extended φ-divergence under the Fundamental Risk Quadrangle framework. The extended φ-divergence is defined for signed measures. We present the primal and dual representations of the quadrangle elements (risk, deviation, regret, error and statistic) and prove that they satisfy the characterizing definitions. The φ-divergence risk measure is studied as a special case of the risk measure. The framework provides an interpretation of portfolio optimization, classification and regression as robust optimization. We furnish illustrative examples demonstrating that many common problems are included in this framework, such as Markowitz portfolio optimization, CVaR optimization, least-squares regression, quantile regression and support vector machine. We conduct a case study to visualize the optimal solution to the inner maximization of robust optimization.

3 - Biased Mean Quadrangle and Applications

Anton Malandii, Stony Brook University, Stony Brook, NY, United States

This paper introduces a novel risk quadrangle called the biased mean quadrangle and examines its mathematical properties. The linear regression using the error function from this quadrangle, referred to as the superexpectation error, can be employed for conditional biased (shifted by a predetermined real number) mean estimation. Minimization of this error is reduced to linear programming. Specifically, when the bias equals zero, the problem of conditional mean estimation reduces to linear programming. This allows for the inclusion of cardinality constraints and reduces the regression problem to a mixed-integer linear program, which is not possible with the standard mean squared error. The reduction to linear programming also indicates the robustness of regression with the superexpectation error to outliers. The theoretical results are validated with case studies.

4 - The Conditional Value-at-Risk (CVaR) based Pickands-type Estimator for Extreme Value Index

Yizhou Li, Stony Brook University, Stony Brook, NY, United States

The Pickands estimator for the extreme value index is advantageous due to its universal consistency and its invariance to location and scale. However, it suffers from poor asymptotic efficiency. To address this, we introduce the Conditional Value-at-Risk (CVaR) into a generalized class of Pickands-type estimators and propose the CVaR-based smoothed Pickands estimator. A simulation study involving a wide range of distributions shows that the CVaR-based smoothed estimator exhibits good finite-sample performance. It outperforms other estimators by achieving a lower mean squared error (MSE) and reduced bias, providing smooth estimates, and maintaining stable performance across intermediate order statistics.

WC34

Summit - 425

Optimized Routing and Travel Time Analysis

Contributed Session

Chair: Rishav Sen

1 - Strategic API Calls: Efficient Arc Travel Time Retrievals for Optimizing a Delivery Route

Mohammad Hesam Rashidi, University of Toronto, Toronto, ON, Canada, Mehdi Nourinejad, Matthew Roorda

Last-mile logistics planners optimize delivery routes primarily based on total travel time. Many, however, lack the extensive historical data necessary for training predictive models to estimate accurate travel times across a network. As a result, they rely on Application Programming Interfaces (APIs) to access travel time data from external sources. APIs have limitations, such as call frequency restrictions and cost per call, which can impact the scalability of data-driven solutions. To address these challenges, this study proposes a heuristic approach that optimizes API usage to determine travel times for solving the Traveling Salesman Problem (TSP). The heuristic starts with an initial TSP solution using Euclidean distances as proxies for actual travel times. This solution is refined through strategic API calls to obtain actual travel times for selected arcs. A Graph Neural Network identifies the most promising arcs to sample based on assessing the local travel time uncertainty across arcs, considering factors like arc length, proximity to arcs with known travel times, and network connectivity. Once actual travel times are sampled, they are extrapolated across the network. The heuristic refines the route using a 2-opt operation—reversing tour segments to reduce overall travel time. This iterative process continues until time or budget constraints are met, ensuring the solution is cost-effective and feasible within real-time operational limits.

2 - Incorporating Travel Time Uncertainty in Last Mile Route Planning Systems

Chhavi Choudhary, Amazon, Hyderabad, India, Zeev Lieber, Yimin Liu, Karthik Konduri

We describe a new approach of informing the route planning systems about the risk of delay using distributional forecast methods for route duration. Amazon's Last Mile operates hundreds of delivery stations worldwide that deliver packages directly to customers' doorsteps. Last Mile's Routing and Planning systems are responsible for creating sustainable routes while maintaining good customer experience. Crucial to that task is the accuracy of route time estimations, which are generated by deep learning models. Route deliveries can sometimes take longer than the planned time due to various factors such as traffic, weather, delivery attempts, and others. The proposed approach allows the planner to directly balance the risk of delay and route efficiency and utilization. Offline and preliminary online results suggest that our approach sacrifices less deliveries per paid hour than the legacy method of reducing risk of delay.

3 - Data-Driven Vehicle Routing Problem with Accident Risk Minimization

Alireza Fallahtafti, Case Western Reserve University, Cleveland, OH, United States, Fatemeh Zandieh, Armin Hamedizad

One of the crucial issues in distributing goods to customers is timely delivery, as any delay in delivering the goods can lead to customer dissatisfaction. Road accidents are among the factors that can significantly impact delivery times. This research presents a data-driven vehicle routing problem with time windows, where the goal is to minimize the vehicle's travel on links with lower accident probabilities. This approach combines optimization of the routing problem with machine learning techniques. In the proposed problem, the objective function considers minimizing accident risks, aiming to increase customer satisfaction by providing timely services.

4 - Shortest Path Problem with a Crash Risk Objective

Qiong Hu, University of Colorado Denver, Denver, CO, United States, Alexander Vinel, Fadel Megahed

With more and more data related to driving, traffic, and road conditions becoming available, there has been renewed interest in predictive modeling of traffic incident risk and corresponding risk factors. New machine learning approaches in particular have recently been proposed, with the goal of forecasting the occurrence of either actual incidents or their surrogates, or estimating driving risk over specific time intervals, road segments, or both. At the same time, as evidenced by our review, prescriptive modeling literature (e.g., routing or truck scheduling) has yet to capitalize on these advancements. Indeed, research into risk-aware modeling for driving is almost entirely focused on hazardous materials transportation (with a very distinct risk profile) and frequently assumes a fixed incident risk per mile driven. We propose a framework for developing data-driven prescriptive optimization models with risk criteria for traditional trucking applications. This approach is combined with a recently developed machine learning model to predict driving risk over a medium-term time horizon (the next 20min to an hour of driving), resulting in a biobjective shortest path problem. We further propose a solution approach based on the k-shortest path algorithm and illustrate how this can be employed.

5 - Grid-Aware Charging and Operational Optimization for Mixed-Fleet Public Transit

Rishav Sen, Vanderbilt University, Nashville, TN, United States, Amutheezan Sivagnanam, Aaron Laszka, Ayan Mukhopadhyay, Abhishek Dubey

Amidst rising urban populations and heightened environmental concerns, transitioning to electric buses has become a pivotal strategy for sustainable urban transit. Managing a mixed fleet of electric, hybrid, and diesel buses presents significant operational challenges, particularly due to the complexities of dynamic electricity pricing. This paper addresses these challenges by developing a comprehensive mixed-integer linear programming (MILP) model that optimizes both the charging schedules and trip assignments for such fleets. Our model distinctively incorporates dynamic electricity pricing and seating constraints, which are critical for adapting to fluctuations in energy costs and meeting specific service demands.

The computational challenges posed by the MILP formulation, typically intractable for large fleets, are tackled using a novel hierarchical solution approach. This approach decomposes the overall problem into more manageable sub-problems that are easier to solve, thereby enhancing the tractability and scalability of the solution. We apply our methodology to real-world data from Chattanooga Area Regional Transportation Authority (CARTA), Tennessee, demonstrating its efficacy. Our results show a 2.58% reduction in operational costs compared to the state-of-the-art model by Sivagnanam et al. (AAAI 2021), a 6.25% improvement over the current operational costs, and a significant reduction (over 800 kg of carbon dioxide daily) in tailpipe emissions currently incurred by the transit agency. Furthermore, our approach aligns with energy efficiency objectives and enhances route service quality by accommodating specific bus capacity requirements.

By systematically addressing economic and environmental factors, our research provides a scalable framework for urban transit agencies seeking to optimize their operations economically and ecologically.

WC35

Summit - 427

Performance Modelling of Intelligent Vehicle Systems

Invited Session

TSL: Intelligent Transportation Systems

Chair: Aadya Bhattarai, University of south florida, Tampa, FL, 33617, United States

1 - Predicting Fuel Consumption for Vehicles with Different Levels of Autonomy Using Data from a Road-trip Simulator

Aadya Bhattarai, University of south florida, Tampa, FL, United States

The rapid integration of autonomy in the transportation sector significantly affects various aspects including fuel consumption, emissions, safety, and road congestion. As vehicles transition through the six levels of autonomy defined by the Society of Automotive Engineers (SAE), they exhibit differences in energy management due to changes in trip planning and execution. This paper addresses a critical gap in the literature by developing models to predict fuel consumption across different levels of autonomy in fossil fuel-driven vehicles. The research leverages a two-phase approach: firstly, constructing a sophisticated simulation model that mirrors real-world driving scenarios; secondly, applying machine learning techniques to predict fuel consumption accurately per mile for various autonomous driving scenarios. The study's findings, derived from simulating realistic vehicle interactions like car following and platooning, are critical for informing policy decisions and enhancing energy optimization in autonomous transportation. This comprehensive analysis not only advances the understanding of fuel dynamics in autonomous vehicles but also sets a foundation for future research in this rapidly evolving field.

2 - Graph-Based Optimization for Efficient Image Segmentation and Object Detection

Aaron Rassiq, University of San Francisco, San Francisco, CA, United States, Deven Varu

Graph theory-based optimization techniques have revolutionized computer vision tasks, particularly in image segmentation and object detection. By modeling pixels and their interrelationships as graphs, these methods enable precise and computationally efficient solutions to challenging problems in computer vision. Key algorithms such as graph cuts and min-cut/max-flow have emerged as powerful tools for segmenting images and detecting objects with high accuracy. Graph-based models capture the spatial dependencies between pixels, allowing for the representation of complex image structures. These models leverage optimization techniques from operations research to partition graphs efficiently, producing high-quality segmentations. The min-cut/max-flow algorithm, for instance, finds optimal boundaries by minimizing the sum of weights on edges that form a cut between the foreground and background regions. Graph cuts further extend this concept to multi-label segmentation, handling multiple object classes in a single framework. This paper provides a novel graph-based optimization method for image segmentation and object detection, illustrating its impact on improving accuracy and efficiency in computer vision applications. Case studies highlight their applications in autonomous driving and real-time surveillance, demonstrating their versatility and effectiveness.

3 - Tacit algorithmic collusion in deep reinforcement learning guided price competition: A study using EV charge pricing game

Diwas Paudel, University of South Florida, Tampa, FL, United States, Tapas Das

Players in pricing games with complex structures are increasingly adopting artificial intelligence (AI) aided learning algorithms to make pricing decisions for maximizing profits. This is raising concern for the antitrust agencies as the practice of using AI may promote tacit algorithmic collusion among otherwise independent players. In this paper, we examine the concern for tacit collusion by considering a practical game where EV charging hubs compete by dynamically varying their prices. The hubs source power from the day-ahead (DA) and real-time (RT) electricity markets as well as from in-house battery storage systems. Their goal is to maximize profits via pricing and efficiently managing the cost of power usage. To aid our examination, we develop a two-step data-driven methodology. The first step obtains the DA commitment by solving a stochastic model. The second step generates the pricing strategies by solving a competitive Markov decision process model using a multi-agent deep reinforcement learning (MADRL) framework. We evaluate the resulting pricing strategies using an index for the level of tacit algorithmic collusion and find a possibility of a low to moderate level of collusion in EV charging market place.

4 - Evaluating Advanced Air Mobility System Performance with An AnyLogic Simulation Tool

Yu Zhang, University of South Florida, Tampa, FL, United States

Early stage use case of passenger advanced air mobility (AAM) could be on-demand and shared station-based services, i.e., passengers accessing to and egressing from vertiports via ground transportation mode(s) and electric vertical take-off and landing aircraft (eVTOL) carrying passengers between vertiports. For providing such services, AAM operators may follow different operational rules, which could lead to different level of service. The authors developed an agent-based AnyLogic simulation tool to mimic the operations of this emerging passenger AAM. For an AAM network designed by the authors' previous study, the simulation tool is used to evaluate system performance with different operational rules.

5 - Managing AV Platoons and Increased AV Penetration on Multiple Lanes of Freeways

Pitu Mirchandani, Arizona State University, Tempe, AZ, United States, Kerem Demirtas, Xuesong Zhou

This study explores the impact of dedicating specific freeway lanes for the exclusive use of Connected and Autonomous Vehicles (AVs) forming platoons, focusing on middle and leftmost lanes. The primary objective is to enhance overall traffic efficiency and safety by structuring and controlling AV movements. Middle lanes are typically less affected by merging and exiting traffic, providing a stable environment for effective platoon formation and maintenance. However, dedicating middle lanes exclusively to AVs can limit access for Human-Driven Vehicles (HDVs) to adjacent lanes, potentially causing congestion. Conversely, the leftmost lane, traditionally used for faster-moving vehicles, offers relative isolation from merging traffic, potentially benefiting AV platoons. Yet, incorporating HDVs in these lanes could lead to speed conflicts and increased lane change complexity. This study compares scenarios where HDVs are either allowed or restricted in dedicated lanes, highlighting the trade-offs between traffic flow flexibility and the stability of AV platoon operations. Allowing HDVs in dedicated lanes offers flexibility but can disrupt AV platoons, whereas restricting HDVs ensures a stable environment for AVs but may cause congestion in non-dedicated lanes. The findings aim to inform traffic managers on controls to optimize lane usage to maximize traffic efficiency and safety.

WC36

Summit - 428

Artificial Intelligence in Logistics/Towards Fair Transportation Planning Under Uncertainty

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Mohammad Moshref-Javadi, University of Illinois at Urbana-Champaign, Champaign, IL, 61820, United States

Co-Chair: Ahmad Hemmati, University of Bergen, Bergen, Norway

Co-Chair: Yu Yang, University of Florida, 401C Weil Hall, Gainesville, 32611-6595, United States

1 - Exploring the Applicability of Machine Learning in Assessing Last Mile Delivery Route Properties

Anna Konovalenko, Molde University College, Molde, Norway, Lars Magnus Hvattum, Mohamed Kais Msakni

Last-mile delivery involves a series of complex tasks in an unpredictable environment. Decision support tools based on optimization algorithms construct efficient routes for drivers, optimizing the cost of making deliveries. However, drivers often deviate from these routes due to factors not considered in the decision-making process. This discrepancy raises the question of how to identify routes that are useable in real-world scenarios. Our research proposes using modern machine learning techniques to classify routes based on their practical usability. In a controlled environment, we demonstrate that machine learning can learn hidden factors influencing route viability by focusing on variants of the vehicle routing problem with additional constraints like time window, capacity and precedence. For each underlying constraint, we show that a machine learning model can be trained to classify routes based on whether or not they violate the constraint. Using datasets generated from well-known benchmark instances, we present computational experiments to evaluate model performance. We discuss which types of constraints are more challenging to recognize and how large a dataset must be to allow for accurate classification. This research has the potential to improve existing decision tools, enabling them to generate routes that better account for real-world complexities.

2 - A Matheuristic for feeder network design problem with optional paired demands and split delivery

Anandhu Dileep, University of Bergen, Bergen, Norway, Norway, Ahmad Hemmati, Rahul Nath, Mohammad Moshref-Javadi

In this work, we study a complex variant of the feeder network design problem from the field of maritime transportation. The problem is about designing routes for ship services between several shipping ports, to deliver cargoes between these ports based on a fixed weekly demand. In many cases, this is done using disjoint networks of ships services, called feeder networks. Each feeder network is connected to a main port by a long ship service called a mother route. Some of the paired demands are delivered by the mother routes, and the remaining demands using daughter routes, which are shorter ship services that branch off from some of the ports visited by the mother routes.

So, the problem is about designing a set of mother routes and associated daughter routes, called feeder networks, to satisfy the paired demands while also allowing split deliveries of the demands and the rejection of some of the demands with penalties. There are time constraints on the routes and also on the travel time for delivery of each cargo and costs involving ship renting, port call and fuel consumption. We develop a multi-stage matheuristic framework to tackle this very complex problem.

We use a large neighborhood search following a Simulated Annealing (SA) paradigm in the heuristic part. A deep reinforcement learning agent is trained for operator selection, based on several different characteristics of the solution and the stage of the search and compared with other approaches for operator selection, like the traditional ALNS.

3 - Congestion Pricing for Efficiency and Equity: Theory and Applications to the San Francisco Bay Area

Chinmay Maheshwari, UC Berkeley, Berkeley, CA, United States

Congestion pricing, while adopted by many cities to alleviate traffic congestion, raises concerns about widening socioeconomic disparities due to its disproportionate impact on low-income travelers. In this study, we address this concern by proposing a new class of congestion pricing schemes that not only minimize congestion levels but also incorporate an equity objective to reduce cost disparities among travelers with different willingness-to-pay. Our analysis builds on a congestion game model with heterogeneous traveler populations. We present four pricing schemes that account for practical considerations, such as the ability to charge differentiated tolls to various traveler populations and the option to toll all or only a subset of edges in the network. We evaluate our pricing schemes in the calibrated freeway network of the San Francisco Bay Area. We demonstrate that the proposed congestion pricing schemes improve both efficiency (in terms of reduced average travel time) and equity (the disparities of travel costs experienced by different populations) compared to the current pricing scheme. Moreover, our pricing schemes also generate a total revenue comparable to the current pricing scheme. Our results further show that pricing schemes charging differentiated prices to traveler populations with varying willingness-to-pay lead to a more equitable distribution of travel costs compared to those that charge a homogeneous price to all.

4 - On the Distributionally Fair Facility Location Problem

QING YE, Georgia Institute of Technology, Atlanta, GA, United States, Weijun Xie

Fairness has been considered in many facility location problems. Since location decisions can have varying impacts on different demographic groups, this paper introduces the concept of distributional fairness to facility location. It aims to minimize distributional disparities of travel costs among groups. We employ the Wasserstein distance to quantify fairness and formulate the problem as a mixed-integer program. Valid inequalities and solution methods are developed to solve the problem effectively.

5 - Fairness Quantification and Assessment in Shared Electric Micromobility Systems

Violet (Xinying) Chen, Stevens Institute of Technology, Hoboken, NJ, United States

Shared Electric Micromobility Systems (SEMSs), such as e-bikes and e-scooters, have been launched in urban centers throughout US. Most SEMSs are burdened with disparities across the service regions and populations. Fair SEMSs are necessary to promote the growth of the micromobility industry and enable multi-modal transportation networks, which are more sustainable and better connected. Motivated by the lack of consistency and comprehensiveness of current approaches to fair SEMSs, we propose a fairness assessment framework to systematically quantify the fairness performance of SEMSs and guide the design of fair geographic distribution of SEMS infrastructure. Our framework strives to be comprehensive and flexible by taking account of different types of impacts SEMSs have on various stakeholders, such as, SEMS users, local communities. Case studies using historical trip data and real-time system feed data from SEMSs in US are developed to demonstrate the proposed framework and draw fairness insights about the current systems.

WC37

Summit - 429

Equity and Optimization in Healthcare Systems

Contributed Session

Chair: Ji-Su Lee, Korea Advanced Institute of Science and Technology, Daejon, Korea, Republic of

1 - Quantifying uncertainties using conformal prediction in predicting catastrophic health expenditure among US households

Behrad Barghi, Rochester Institute of Technology, Rochester, NY, United States, Nasibeh Azadeh-Fard

Catastrophic health expenditure (CHE) describes situations where a family's direct medical expenses are so extensive that they lead to financial struggles or necessitate significant cutbacks in other vital areas, like food or housing. Health economics identifies two primary methods for estimating CHE: budget share (BS) and capacity-to-pay (CTP). This study utilizes the medical expenditure panel survey (MEPS) to explore CHE in US households. Since MEPS omits food spending data, this research uses the budget share approach. The study applies three benchmarks: 10%, 15%, and 20%, to determine CHE. Tools like the concentration index (CI), Wagstaff index (WI), and Erreygers Index (EI) were employed to measure and break down socioeconomic disparities in CHE. Additionally, this study employed 5 machine learning techniques, including random forest (RF), artificial neural networks (ANN), xgboost (XG), support vector machine (SVM), and deep learning (DL) to find the best models to forecast, categorize, and pinpoint the key factors of CHE during the specified timeframe. Conformal prediction is a technique used to quantify the uncertainty of predictions made by machine learning models. This method is useful for assessing the reliability of predictions and ensuring that the models used can provide robust and accurate forecasts. In this paper, conformal prediction method is used to find the best significance level and robust model to predict risk of CHE on US households.

2 - Closing the Loop of Healthcare Quality Improvement: An Empirical Study of Quality Competencies

Xinyu Wei, California State University, Chico, Chico, CA, United States, Heng (John) Xie, Richard Peng, Victor Prybutok

Assessment of healthcare quality is vital for enhancing patient outcomes and achieving organizational excellence. As such, assessment is of interest to healthcare professionals and scholars. This research investigates an industry framework of healthcare quality competencies to explore relevant measurements arising from the dynamic nature of today’s healthcare quality landscape. We integrate healthcare quality practices and results to develop a research model and validate its effectiveness using survey data obtained from healthcare workers. Findings are discussed to provide insights to practitioners and researchers in the healthcare field.

3 - A Study on Quality Factors for Continuous Use of Digital Healthcare Services

Seonghyeon Han, Korea University, Seoul, Korea, Republic of, Jaeseok Yoon, Sung Yeon Kim, Jinmin Kim

The digitalized networking society is increasing the scope and quality of medical services while improving access to health information and public health education, and as a result, digital healthcare services are expected to be a high value-added service industry, showing a lot of expectations and demand. However, looking at the existing research, there were many studies claiming the discrepancy between the expected benefits and actual implementation related to digital healthcare services, and few studies were found for the continuous use of the service, focusing on the effect of technology adoption and user reliability and experience on service use. Therefore, this study aims to derive quality factors that affect the continuous use of digital healthcare services by analyzing the quality of products and services that make up digital healthcare services. This study adopted the VAM (Value-based Adoption Model) to construct a research model, and set the personal characteristics claimed in the previous study and degree of standardization as control variables to suggest quality factors to be considered when establishing a service standardization strategy for continuous use. This study is expected to provide basic data necessary for service standardization and service certification by separating and analyzing device and service quality for the continuous use of digital healthcare services, which are expected to be high value-added services, and providing priority for service quality factors by reflecting changes and prospects in the service industry.

4 - A Game Theoretic Approach to Efficient and Equitable Healthcare Quality Improvement Policies for Regional Disparity

Ji-Su Lee, KAIST, Daejeon, Korea, Republic of, Woonam Hwang, Taesik Lee

While improving the quality of healthcare is a paramount objective for any healthcare systems, regional disparities in quality must not be overlooked to preserve equity and fairness in healthcare. In this respect, South Korea faces a distinctive challenge, wherein patients residing in non-metropolitan regions often gravitate towards major hospitals located in the metropolitan area, driven by perceptions of superior quality in those facilities.

To explain this phenomenon, this study constructs a modified Salop model in which hospitals under a price regulation compete on quality to attract patients seeking maximum utility. Our game theoretic analysis reveals an equilibrium where hospitals in the metropolitan region attain a higher level of quality compared to other hospitals. This equilibrium stems from two factors: intensified competition within the metropolitan area and disparities in the respective catchment population sizes.

We then use this model to explore policy options aimed at efficiently boosting overall healthcare quality while reducing regional disparities. Under this goal, a desirable policy would disproportionately enhance the quality of underperforming rural hospitals. We cast this problem as a Stackelberg game, where the government first sets a policy, followed by hospitals determining their quality levels, and patients subsequently choosing which hospital to visit. The specific policies evaluated include fee-for-service supplemental payments, pay-for-performance schemes linking reimbursement to quality metrics, and performance-based subsidies to hospitals. Equilibrium solutions resulting from each policy are analyzed to identify the optimal approach for achieving high-quality care while promoting regional equity.

WC38

Summit - 430

Advances in Vehicle Routing

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: John Carlsson, University of Southern California, Los Angeles, CA, United States

1 - Demand Equilibria in Spatial Service Systems

John Carlsson, University of Southern California, Los Angeles, CA, United States

A service is offered at certain locations ("facilities") in a geographical region. Customers can appear anywhere in the region, and each customer chooses a facility based on travel distance as well as expected waiting time. Customer decisions affect waiting times by increasing the load on a facility, and thus impact other customers' decisions. The service provider can also influence service quality by adjusting service rates at each facility. Using a combination of queueing models and computational geometry, we characterize demand equilibria in such spatial service systems. An equilibrium can be visualized as a partition of the region into service zones that form as a result of customer decisions. Service rates can be set in a way that achieves the best possible social welfare purely through decentralized customer behavior.

2 - Penalty functions for truck-and-drone routing

Ke Xu, University of Southern California, Los Angeles, CA, United States

This study studies efficient last-mile delivery using trucks and drones. We introduce the Pickup-and-Delivery Generalized Traveling Salesman Problem (PDGTSP), a new truck-and-drone routing variant. Our approach uses a heuristic method based on Lin-Kernighan-Helsgaun with penalty functions to solve this problem. Computational experiments validate our method's effectiveness. We also discuss potential applications to more complex scenarios like the Traveling Salesman Problem with Multiple Drones (TSP-MD) and the Vehicle Routing Problem with Drones (VRPD).

3 - Sheng Liu's (student's) session

Sheng Liu, U of Toronto, Toronto, ON, Canada

4 - A continuous approximation model for the close-enough TSP

Mingxi Wang, University of Southern California, Los Angeles, CA, United States

We give a continuous approximation analysis of the “close-enough TSP”, in which one is given a set of shapes in a service region and the goal is to find the shortest tour that visits at least one point in each region. Our analysis focuses on the case where the regions are disks in the Euclidean plane. We derive asymptotic results for the optimal tour length as the number of regions approaches infinity, while the total area covered by the regions remains approximately constant. Using techniques from geometric probability theory, we establish tight bounds on the tour length.

WC39

Summit - 431

Strategic Behavior in Queueing and Service Systems

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Chen-An Lin, Purdue University, West Lafayette, IN, 47906, United States

1 - Adding Servers to Reinforce Facilities in A Spatially Distributed Queueing System

Monir Sabbaghtorkan, UCLA, Los Angeles, CA, United States

We study the problem of adding servers to reinforce functioning facilities modeled as spatially distributed M/M/c type queuing systems. We consider two modes, stationary and interactive. The stationary mode is for a situation in which adding new servers to functioning facilities will not change the demand distribution (e.g., hospital care or vehicle inspection). The interactive mode is for a situation in which adding servers will likely change the demand distribution (e.g., gas stations, electric vehicle charging stations). Here we use the gravity rule to model demand adjustment. A case study is presented based on the allocation of gas servers to existing gas stations in a hurricane evacuation event. Managerial insights are provided based on sensitivity analysis.

2 - Waiting Together: How to Infer Response TIME when Customers Wait in Groups

Yanting Li, University of Rochester, Rochester, NY, United States, Ricky Roet-Green

We introduce a novel model that accommodates customers arriving either individually or in pairs, who are then served in their original group configurations. Each group is demanding one unit of service, regardless of the group size. The concept of information granularity is central to our study, indicating that upon arrival, customers may possess incomplete information about the demand distribution. To analyze customer strategies in the absence of perfect information, we employ Bayesian updating process to capture customers' decision making. Through an evaluation of the value of demand ratio information, we show that the provision of additional information does not always yield optimal throughput or social welfare. Indeed, under specific conditions, less information may enhance system performances. Consequently, service providers should carefully determine the level of information sharing, taking into consideration the parameters of the system.

3 - Optimal Control of Single-Server Queues: a Wait-TIME-Based Approach

Chen-An Lin, Purdue University, West Lafayette, IN, United States

This paper studies an optimal control problem for a single-server queue in which customers arrive according to a Poisson process. The service provider determines a control policy for each arriving customer according to the system wait time. The customer, in turn, decides whether to accept the policy offer and join the queue. The objective is to maximize the average reward in an infinite horizon. The proposed model encompasses many practical service systems with customers ranging from homogeneous with a known type to heterogeneous with unknown types. We illustrate the applicability of this general model through three practical service systems: discretionary services, service operations, and make-to-order systems. The key contribution of this work is to show the existence and optimality of a stationary, wait-time-dependent policy. In addition, we characterize the structure of the optimal policy through the corresponding Hamilton-Jacobi-Bellman equation and reveal insights. The application of the make-to-order system leads to a mechanism design problem. In particular, we find that when the wait time becomes longer, the provider should admit more patient customers while offering an option with a shorter job lead time. Interestingly, the offered payment may not be monotonic in the wait time.

4 - Managing Service Systems with Overconfident Customers

Na Zhang, Wichita State University, Wichita, KS, United States

We study a service system where true service times are the manager's private information and customers exhibit the cognitive bias of being overconfident (in particular, overprecise) in their beliefs about service times—customers underestimate the variability of service times. We formulate the problem as a stylized queueing model to examine the implications of overconfidence in service systems. Our models and results generalize those in several seminal queueing-economics papers published on Econometrica. Specifically, when customer overconfidence level is zero, our models and results reduce to those in the aforementioned papers; however, when customer overconfidence level is positive, our models and results are qualitatively different. Our results also generate interesting practical implications. First, we provide overconfidence theory as a plausible explanation for the commonly observed long-queue phenomenon. Second, the manager should adopt a strictly increasing price patter w.r.t. overconfidence level in the unobservable queue, but he may need to implement a non-monotonic price pattern in the observable queue. Third, we recommend the manager to always reveal queue length information to improve revenue for both sufficiently low and sufficiently high congestion, but strategically decide whether to reveal queue length information for intermediate congestion. Additionally, customers should be wary of the ostensible benefit of queue length information as it may decrease consumer surplus. This paper unravels the role of overconfidence in service systems and its important implications on the manager's pricing decision and queue-length-information provision policy, as well as consumer surplus.

WC40

Summit - 432

Data-Driven Healthcare Management

Contributed Session

Chair: Shirin Geranmayeh, University of Alberta, Edmonton, AB, Canada

1 - Two-Step Machine Learning Approach in Healthcare Fraud Detection

Basiru Usman, North Carolina State University, Raleigh, NC, United States

In the realm of healthcare, a multifaceted environment intertwines with political sensitivity, necessitating meticulous scrutiny of any occurrences within the industry. Concurrently, the sector grapples with escalating data volumes, fostering an environment ripe for fraudulent activities exploiting its intricacies and sensitivities. This study introduces a two-tiered machine learning framework designed to augment healthcare fraud auditors' decision-making processes, offering invaluable support in navigating this complex landscape.

2 - Physician-Patient Co-production of Healthcare Services: Leveraging Generative Artificial Intelligence (AI) for Smarter Decisions and Value Co-creation

Arvin Sahaym, Washington State University, Pullman, WA, United States, Om Sahaym, Jay Sahaym

A healthcare ecosystem is a complex and dynamic system involving multiple stakeholders. Assuming that patients in the modern era get information (or misinformation) about their symptoms on the internet, this study contributes by urging traditional healthcare ecosystems to evolve into a system of co-production rather than remaining stuck in the tradition of 'service delivery.' Building on the premise that the overall goal is to provide superior value in terms of patient-centered care *and* patient satisfaction, this study examines how patients and specialist physicians can co-produce healthcare leveraging AI for smarter decisions and value co-creation. We draw on the patient's voice and feedforward literature, decision-making, and behavioral economics to propose that 1) physicians' proactive consideration of patients' voice in clinical decisions has costs and benefits but will be associated with net gain in providing superior value, 2) patient voice is influenced by generative AI, which provides more detailed deep-learning-based information that will further enhance the above relationship, 3) the AI-influenced patient voice nudges physicians toward considering the treatment desired by the patient, and 4) AI usage increases psychological ownership of the treatment by the patient, leading to higher levels of patient-centered care *and* satisfaction. Overall, smarter decision-making and superior value will be attained by incorporating AI-enabled patient voice. We contribute to the literature on decision-making, psychology, and economics and have implications for healthcare stakeholders and policymakers.

3 - Managing the Gatekeeping System: Referral Discounts and AI-Based Diagnosis

Chenhe Pan, Renmin University of China, Beijing, China, People's Republic of, Xiaofang Wang, Yixuan Liu, Guoming Lai

Patients may suffer from high congestion and delayed treatment due to unevenly distributed healthcare resources. Without professional assistance, patients cannot accurately identify their severity and may make inappropriate choice among tiered care providers. Recently, artificial intelligence-based diagnosis is widely used to help patients to make better treatment choice in addition to traditional financial incentives. We develop a model of the two-level gatekeeping healthcare system with patient choice to study the referral incentive design and the impact of introducing such artificial intelligence-based diagnosis. Our results provide insights from the perspective of social welfare optimization.

4 - Clinical Trials Supply Chain: a Multi-Stage Stochastic Programming Approach for Inventory Control

Mustafa Gurbuz, Zaragoza Logistics Center, ZARAGOZA, Spain, Daniel Calcinaro

Clinical trial supply chains have recently become more complex due to increasing globalization of such experiments and strict regulations in the pharmaceutical industry. Managers are under pressure to secure early entry to the market (i.e., reduce completion time of the trial), reduce supply chain costs, and reduce excess inventory at the end of the trial.

We propose a multi-stage stochastic optimization model to make informed production and distribution decisions with the concerns mentioned above in mind. The clinical trial supply chain we analyze has one central warehouse, multiple country depots, and multiple sites (e.g., hospitals). Patients enrol for the new drug at the sites according to a known probability distribution. Clinical trial ends when a targeted number of patients enrol, therefore we analyze a finite horizon model without a specified completion time. Our model is different from most of the existing models in the relevant literature because it allows for multiple production runs, multiple dose, and the perishable nature of the drugs.

As expected, the optimization model requires a prohibitively large number of scenarios to be evaluated. Therefore, a novel clustering-based scenario reduction technique is proposed to quickly obtain near optimal solutions. We test our model in three clinical trials that a global major pharmaceutical company ran in recent years. The scenario reduction technique we propose works very well, and show that our model helps minimize completion time, while keeping costs under control and reducing excess inventory.

5 - Operating room planning and modelling of a network of post-op units in a paediatric tertiary teaching hospital

Shirin Geranmayeh, University of Alberta, Edmonton, AB, Canada, Armann Ingolfsson, John Doucette, Oleksandr Shlakhter

Our research explores methodologies for optimizing operating room (OR) scheduling at the Stollery Children’s Hospital - a pediatric tertiary teaching hospital in Edmonton, Alberta, Canada. We propose an approach aimed at developing a master surgery schedule to minimize post-operative bed blockages in downstream units specific to a pediatric hospital. The objective is to address the challenges of surgery cancellations due to bed shortages and the aim is to improve resource allocation efficiency. Historical data is used to estimate the patient-age-specific consumption of hospital resources, which impacts operational decision making in OR, and, bed allocation and utilization.

WC41

Summit - 433

Information Design for Incentive-Aware Learning

Invited Session

Applied Probability Society

Chair: Kunhe Yang, UC Berkeley, Berkeley, CA, United States

Co-Chair: Nika Haghtalab

1 - Information Design for Collaborative Learning: Efficiency, Incentive-Awareness, and Robustness

Kunhe Yang, University of California, Berkeley, Berkeley, CA, United States, Nika Haghtalab, Mingda Qiao

We explore an information design perspective on promoting collaboration among strategic agents, with applications to collaborative learning. It has been shown that many traditional collaboration protocols either require efforts from certain agents that surpass what is individually rational for them, or suffer from significant inefficiency. Motivated by federated collaborative learning, we propose a framework where an informed platform that possesses more information about the relationship between agents' tasks aim to steer the agents towards collaboration outcomes that are both efficient and incentive-aware. We design information structures that achieve efficiency, incentive-awareness, and robustness to signal leakage at the same time.

2 - A Persuasive Approach to Combating Misinformation

Safwan Hossain, Harvard University, Cambridge, MA, United States, Andjela Mladenovic, Yiling Chen, Gauthier Gidel

Bayesian Persuasion is proposed as a tool for social media platforms to combat the spread of misinformation. Since platforms can use machine learning to predict the popularity and misinformation features of to-be-shared posts, and users are largely motivated to share popular content, platforms can strategically signal this informational advantage to change user beliefs and persuade them not to share misinformation. We characterize the optimal signaling scheme and utility when predictions are imperfect by framing this as a linear program and giving sufficient and necessary conditions on the classifier to ensure optimal platform utility is non-decreasing and continuous. Next, this interaction is considered under a performative model, wherein platform intervention affects the user's future behaviour. The convergence and stability of optimal signaling under this performative process are fully characterized. We also experimentally validate that our approach leads to a significant reduction in misinformation, even with weak classifiers, and comment on the broader scope of using information design to combat misinformation.

3 - Incentivizing Combinatorial Bandit Exploration

Daniel Ngo, University of Minnesota, Pittsburgh, PA, United States, Xinyan Hu, Alex Slivkins, Steven Wu

Consider a bandit algorithm that recommends actions to self-interested users in a recommendation system. The users are free to choose other actions and need to be incentivized to follow the algorithm's recommendations. While the users prefer to exploit, the algorithm can incentivize them to explore by leveraging the information collected from the previous users. All published work on this problem, known as incentivized exploration, focuses on small, unstructured action sets and mainly targets the case when the users' beliefs are independent across actions. However, realistic exploration problems often feature large, structured action sets and highly correlated beliefs. We focus on a paradigmatic exploration problem with structure: combinatorial semi-bandits. We prove that Thompson Sampling, when applied to combinatorial semi-bandits, is incentive-compatible when initialized with a sufficient number of samples of each arm (where this number is determined in advance by the Bayesian prior). Moreover, we design incentive-compatible algorithms for collecting the initial samples.

4 - Dynamic Pricing and Learning with Bayesian Persuasion

Shipra Agrawal, Columbia University, New York, NY, United States

We consider a novel dynamic pricing and learning setting where in addition to setting prices of products in sequential rounds, the seller also ex-ante commits to 'advertising schemes'. That is, in the beginning of each round the seller can decide what kind of signal they will provide to the buyer about the product's quality upon realization. Using the popular Bayesian persuasion framework to model the effect of these signals on the buyers' valuation and purchase responses, we formulate the problem of finding an optimal design of the advertising scheme along with a pricing scheme that maximizes the seller's expected revenue. Without any apriori knowledge of the buyers' demand function, our goal is to design an online algorithm that can use past purchase responses to adaptively learn the optimal pricing and advertising strategy. We study the regret of the algorithm when compared to the optimal clairvoyant price and advertising scheme. Our main result is a computationally efficient online algorithm that achieves an $O(T^{2/3}(m\log T)^{1/3})$ regret bound when the valuation function is linear in the product quality. Here m is the cardinality of the discrete product quality domain and T is the time horizon. This result requires some natural monotonicity and Lipschitz assumptions on the valuation function, but no Lipschitz or smoothness assumption on the buyers' demand function. For constant m, our result matches the regret lower bound for dynamic pricing within logarithmic factors, which is a special case of our problem. We also obtain several improved results for the widely considered special case of additive valuations, including an $\tilde{O}(T^{2/3})$ regret bound independent of m when $m\le T^{1/3}$.

5 - Bayesian Strategic Classification

Juba Ziani, Georgia Tech, Atlanta, GA, United States

In strategic classification, agents modify their features, at a cost, to obtain a positive classification from the learner's classifier. The typical response of the learner is to carefully modify their classifier to be robust to such strategic behavior. Most works rely on the strong assumption that agents fully know the exact parameters of the deployed classifier by the learner. This can be unrealistic when using complex or proprietary machine learning techniques in real-world prediction tasks.

We initiate the study of partial information release by the learner in strategic classification. We move away from the traditional assumption that agents have full knowledge of the classifier and consider agents that have a distributional prior on which classifier the learner is using. The learner can reveal truthful, yet not necessarily complete, information about the deployed classifier to the agents. The learner's goal is to release just enough information about the classifier to maximize accuracy. We show how such partial information release can, counter-intuitively, benefit the learner's accuracy, despite increasing agents' abilities to manipulate.

We show that while it is intractable to compute the best response of an agent in the general case, there exist oracle-efficient algorithms that can solve this best response problem for the class of linear classifiers, or when the agents' cost function satisfies a natural notion of submodularity. We then turn our attention to the learner's optimization problem and provide both positive and negative results on the algorithmic problem of how much information the learner should release about the classifier.

WC42

Summit - 434

Theoretical Advances in Networks, Dynamics and Inference

Invited Session

Applied Probability Society

Chair: Anirudh Sridhar, MIT, Cambridge, MA, United States

1 - Exact Random Graph Matching with Multiple Graphs

Taha Ameen, University of Illinois Urbana-Champaign, Urbana, IL, United States, Bruce Hajek

This work studies fundamental limits for recovering the underlying correspondence among multiple correlated random graphs. We identify a necessary condition for any algorithm to correctly match all nodes across all graphs, and propose two algorithms for which the same condition is also sufficient. The first algorithm employs global information to simultaneously match all the graphs, whereas the second algorithm first partially matches the graphs pairwise and then combines the partial matchings by transitivity. Remarkably, both algorithms work down to the information theoretic threshold. Our analysis reveals a scenario where exact matching between two graphs alone is impossible, but leveraging more than two graphs allows exact matching among all the graphs. Along the way, we derive independent results about the k-core of Erdos-Renyi graphs.

2 - Interplay of Network Structure and Evolutions in the Emergence of Epidemics

Mansi Sood, Carnegie Mellon University, Pittsburgh, PA, United States, Anirudh Sridhar, Rashad Eletreby, Chai Wah Wu, Simon A Levin, H. Vincent Poor, Osman Yagan

In this talk, we examine mechanisms that lead to the widespread propagation of contagions, e.g., misinformation and pathogens, and identify risk factors that can trigger widespread outbreaks. Akin to different strains of a pathogen arising through evolutionary adaptations or mutations, different variants of a piece of information may be created as the content is altered on social media platforms. The resulting variants of the information may have different propensities to be circulated on different social media platforms. However, most existing models for analyzing contagions assume a simplistic network structure with a homogeneous probability of transmission of the contagion across different types of links and contagion strains. We propose and analyze models for the spread of evolving contagions by accounting for a multi-layer network structure where each individual can participate in different network layers corresponding to different settings in which they interact with others and potentially pass on the contagion. Our results highlight that reductions to existing models that do not simultaneously account for heterogeneity in the contagion strains and network layers may lead to incorrect predictions of the likelihood of the emergence of an epidemic outbreak.

3 - Random Geometric Graph Alignment with Graph Neural Networks

Suqi Liu, Harvard University, Boston, MA, United States

We characterize the performance of graph neural networks for graph alignment problems in the presence of vertex feature information. More specifically, given two graphs that are independent perturbations of a single random geometric graph with noisy sparse features, the task is to recover an unknown one-to-one mapping between the vertices of the two graphs. We show under certain conditions on the sparsity and noise level of the feature vectors, a carefully designed one-layer graph neural network can with high probability recover the correct alignment between the vertices with the help of the graph structure. We also prove that our conditions on the noise level are tight up to logarithmic factors. Finally we compare the performance of the graph neural network to directly solving an assignment problem on the noisy vertex features. We demonstrate that when the noise level is at least constant this direct matching fails to have perfect recovery while the graph neural network can tolerate noise level growing as fast as a power of the size of the graph.

4 - Manifold Filters and Neural Networks: Geometric Graph Signal Processing in the Limit

Zhiyang Wang, University of Pennsylvania, Philadelphia, PA, United States, Luana Ruiz, Alejandro Ribeiro

Convolutional neural networks (CNNs) have achieved impressive success in a wide range of applications. When processing non-Euclidean data, graphs are commonly used as a discrete model to capture the underlying geometric structure. Convolutions can be readily extended to graph convolutions, which allows defining convolutional graph neural networks (GNNs). In graphs of moderate size, GNNs are well-backed by their expressive power and stability. However, in the large-scale regime with more practical interest, their behavior is not as well understood. My work focuses on building manifold convolutional filters and manifold neural networks (MNNs) as a limit model for convolutional filters and neural networks on large geometric graphs. A two-way connection between filters and neural networks on graphs and manifolds is proposed by showing GNNs running on graphs sampled from a manifold converge to the MNN running on the underlying manifold with non-asymptotic convergence results. The convergence allows the transferability of GNNs among different graphs sampled from the same underlying manifold, which enables us to design GNNs on small graphs and transfer them to larger graphs. MNNs, as a formal limit of GNNs provide a theoretical tool for understanding large-scale GNNs. I proceed to analyze the stability of manifold filters and MNNs to smooth deformations of the manifold. This sheds light on the behavior of large-scale graph filters and GNNs, which are prone to undergo perturbations and changes in real-world scenarios. The theoretical results are verified under wireless resource allocation, point cloud analysis, and navigation control problem settings.

WC43

Summit - 435

Advances in Deep Learning Theory

Invited Session

Applied Probability Society

Chair: Boix-Adserà Enric, Cambridge, MA, United States

1 - How do neural networks learn features from data?

Adityanarayanan Radhakrishnan, Harvard SEAS, Cambridge, MA, United States, Daniel Beaglehole, Parthe Pandit, Mikhail Belkin

Understanding how neural networks learn features, or relevant patterns in data, for prediction is necessary for their reliable use in technological and scientific applications. We propose a unifying mechanism that characterizes feature learning in neural network architectures. Namely, we show that features learned by neural networks are captured by a statistical operator known as the average gradient outer product (AGOP). Empirically, we show that the AGOP captures features across a broad class of network architectures including convolutional networks and large language models. Moreover, we use AGOP to enable feature learning in general machine learning models through an algorithm we call Recursive Feature Machine (RFM). Overall, this line of work advances our fundamental understanding of how neural networks extract features from data, leading to the development of novel, interpretable, and effective models for use in scientific applications.

2 - Moreau Envelopes, Overfitting, and Generalization Theory

Frederic Koehler, University of Chicago, Chicago, IL, United States

Memorization of the training data, in various forms, seems to be a common feature of many successful ML models. From a theoretical standpoint, understanding the interplay between partial memorization and generalization also turns out to be very interesting. We discuss some recent developments which have connected VC-style generalization theory to modern probabilistic methods in high-dimensional statistics and given some new insight into what's going on, especially in the benchmark setting of linear models.

3 - The Neural Covariance SDE and Its Limiting Spectrum

Mufan Li, Princeton University, Princeton, NJ, United States

Recent advances in neural network research have predominantly focused on infinite-width architectures, yet the complexities inherent in modelling networks with substantial depth call for a novel theoretical framework. In this presentation, we explore a unique approach to modelling neural networks using the proportional infinite-depth-and-width limit.

In fact, naively stacking non-linearities in deep networks leads to problematic degenerate behaviour at initialization. To address this challenge and achieve a well-behaved infinite-depth limit, we introduce a fundamentally novel framework: we treat neural networks as depthwise stochastic processes. Within this framework, the limit is characterized by a stochastic differential equation (SDE) that governs the feature covariance matrix. Notably, the framework we introduced leads to a very accurate model of finite size networks, and the approach yields a natural remedy for vanishing gradients in Transformers. Finally, we will also leverage the SDE structure to derive a limit spectrum result for feature covariance of linear networks.

4 - How Transformers Learn Causal Structure with Gradient Descent

Eshaan Nichani, Princeton University, Princeton, NJ, United States

The incredible success of transformers on sequence modeling tasks can be largely attributed to the self-attention mechanism, which allows information to be transferred between different parts of a sequence. Self-attention allows transformers to encode causal structure which makes them particularly suitable for sequence modeling. However, the process by which transformers learn such causal structure via gradient-based training algorithms remains poorly understood. To better understand this process, we introduce an in-context learning task that requires learning latent causal structure. We prove that gradient descent on a simplified two-layer transformer learns to solve this task by encoding the latent causal graph in the first attention layer. The key insight of our proof is that the gradient of the attention matrix encodes the mutual information between tokens. As a consequence of the data processing inequality, the largest entries of this gradient correspond to edges in the latent causal graph. As a special case, when the sequences are generated from in-context Markov chains, we prove that transformers learn an induction head (Olsson et al., 2022). We confirm our theoretical findings by showing that transformers trained on our in-context learning task are able to recover a wide variety of causal structures.

WC44

Summit - 436

Business Strategy

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Qilong Zhu, Texas A&M University, COLLEGE STATION, United States

Co-Chair: Qilong Zhu, Texas A&M University, COLLEGE STATION, United States

1 - A Data-Driven Optimization Framework to Improve Search Advertising Efficiency

Kerim U. Kizil, Texas A&M University, College Station, TX, United States, Rajiv Mukherjee, Bala Shetty, Chelliah Sriskandarajah, Jon Stauffer

This study proposes a data-driven optimization framework to improve search advertising efficiency in higher education marketing. We define efficiency in this context as the number of clicks per advertisement dollar spent. Using a search term dataset provided by the search engine, we investigate the case of a specific higher education institution. Social media and search term data are analyzed with natural language processing and AI tools to identify seasonal patterns in the decision-making journeys of prospective students. Based on marketing literature, we focus on four decision-making steps: *(i)* need recognition, *(ii)* information search, *(iii)* evaluation of alternatives, and *(iv)* decision/application. We subsequently categorize the search terms in our dataset into four classes: *(i)* explorative, *(ii)* navigational, *(iii)* rival-branded, and *(iv)* focal-branded. The built empirical models inform the optimization process of keyword-level bid amounts, which are subject to change temporally. We use the results of this study to argue that advertisers can achieve higher search advertising efficiency by analyzing available data (i.e., historical search terms and social media data) using machine learning and AI methods rather than relying on the proprietary and opaque algorithms of search engines for bid decisions.

2 - Strategic Product Design and Launch Strategies for Manufacturers When Selling to Time-Inconsistent Consumers

Chen Pang, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Gang Li, Li Jiang

The temporal disparity between instant payments and delayed payoffs leads to the misestimation of intertemporal utilities, resulting in consumers' time-inconsistent behaviour. This paper explores how manufacturers strategically adjust their product design and product launch strategies if consumers are time-inconsistent. We develop a two-period utility-based model by employing a hyperbolic discounting scheme. Our results reveal that time inconsistency gives rise to three demand patterns to the manufacturer: demand vanishing, demand migration, and demand expansion. The extent of time inconsistency and the differentiation level between advanced and original versions interact dynamically, influencing the magnitude of demand patterns accordingly. Moreover, when the advanced-versioned product closely resembles the original-versioned product in experiential features, the manufacturer tends to accelerate its product launch schedule in the presence of time-inconsistent consumers, a phenomenon termed the assistance effect. Conversely, a deterrent effect is observed when the experiential features of the advanced-versioned product significantly diverge from those of the original-versioned product. In such case, the manufacturer is more likely to defer the implementation of the product launch strategy when time inconsistency exists.

3 - Balancing Act: The impact of ESG disclosure on customer network instability

JIA REN, The Hong Kong Polytechnic University , Hong Kong, Hong Kong, Kee-hung Lai

Environmental, social, and governance (ESG) disclosure is gaining stakeholder attention.A firm's operations may be adversely affected by excessive or insufficient disclosure of ESG, a non-financial objective. So ESG disclosure balance deserves more attention. Besides, the scope of a firm's ESG disclosure is crucial for its customers because customers are the receiving ends of ESG practices and then they will decide on whether to exit original supply and demand network based on their upstream firms' ESG disclosure. However, the impact of a firm's ESG disclosure balance on its customers network instability remains a question. Drawing on organizational ambidexterity theory, this research finds an instability-reducing effect of ESG disclosure balance on a firm's customer network from the perspective of exploitation. Using 2012-2022 panel dataset from multiple sources, we find that ESG disclosure balance is negatively related to customer network instability. Furthermore, this study identifies two boundary conditions affecting this relationship.Specifically, AI innovation, as an exploration activity, strengthens the relationship between ESG disclosure balance and customer network instability, whereas from the perspective of financial constraints, positive feedback on financial performance weakens the main effect. These findings contribute to ESG disclosure and sustainable supply chain management literature by illustrating the effect of a firm's ESG disclosure balance on its customer network instability. In addition, by elaborating on exploitation (ESG disclosure balance) and exploratory activities (AI innovation), this study contributes to organizational ambidexterity theory advancement and expansion in the context of ESG disclosure.

4 - Investigating community equity through franchised fast-food restaurants: A crowdsourcing approach using Google Map Reviews

Hongli Ye, Clemson University, CENTRAL, SC, United States, Lingyao Li, Songhua Wu, Weiwei Zhan

Franchised restaurants are expected to maintain uniformity across the nation. However, through an analysis of Google Maps reviews using large language models (LLMs) and sentiment analysis, we uncovered significant variations in service quality across different communities. Furthermore, our findings indicate that the COVID-19 pandemic has notably influenced community equity within the franchised restaurant sector.

WC45

Summit - 437

From Data to Action: Analytics in Public Health

Invited Session

Health Applications Society

Chair: Ebru Bish, University of Alabama, Tuscaloosa, AL, United States

Co-Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, 95053, United States

1 - Data-Informed Changes in Prenatal Care Policies to Meet Social as Well as Medical Needs

Amy Cohn, University Of Michigan, Ann Arbor, MI, United States

The standard practice of prenatal care in the United States has seen minimal changes over the history of modern healthcare. It is arguably the antithesis of "personalized medicine," with standard pathways of in-person clinical visits with limited deviation. Arguably, these visits may not all be value-added, especially when taking into account the fact that in many parts of the country the healthcare system is at (or beyond) capacity. Furthermore, while lower-income patients often receive too little care in the US health system, arguably the modern-day prenatal pathways may provide too *much* care. For patients with limited income, transportation access, and job flexibility, additional prenatal visits can provide a high burden without significant added value. We use data-driven simulation to evaluate the impact of different prenatal care pathways on patient access and clinic capacity.

2 - Cost-Effectiveness Analysis of Reducing Exposure to Secondhand Smoke in New York City: A Cardiovascular Disease Microsimulation Study

You Zhou, Purdue University, West Lafayette, IN, United States, Nan Kong, Brandon Bellows, Yan Li

Cardiovascular disease is the leading cause of deaths in the US. To reduce its heavy burden on the healthcare system, prevention-based public health policies become the focus through promoting healthy behaviors and managing risk factors. In our paper, we adapted the well-established and validated microsimulation model for tobacco tax policy analysis in New York city with incorporation of 2 key factors: 1) illicit market and 2) secondhand smoke exposure. We simulated each individual's cardiovascular disease (e.g., stroke and coronary heart disease) progression trajectory, and estimated the associated health cost reduction and the quality-adjusted life years (QALYs) gained by implementing tax policy to reduce cigarette consumption and secondhand smoke exposure for adults in New York city. In the model, the population socioeconomical and health profiles were sampled from NYC Health and Nutrition Examination Survey dataset and the relative risk of secondhand smoke exposure on cardiovascular disease were pooled from the recent meta-analyses. Our preliminary results show that the additional tobacco tax can be cost-effective (defined as <$50,000 per QALY) even with substantial illicit market. Health benefits from nonsmokers exposed to secondhand smoke show significant contribution in the analysis which takes up over 27% of total incremental QALYs.

3 - Optimizing Surveillance Policies for Hospital-Acquired Infections: Case of Mrsa

Esma Akgun, University of Waterloo, Waterloo, ON, Canada, F. Safa Erenay, Sibel Alumur Alev, William Ciccotelli

Roommates of nosocomial methicillin-resistant \textit{Staphylococcus} aureus (MRSA) cases have a high risk of MRSA acquisition. Following infection prevention and control guidelines, these individuals are isolated and undergo surveillance screening. However, the optimal post-exposure surveillance testing and isolation strategies for contacts of MRSA cases are unknown. We develop a Markov decision process model to optimize the screening decisions for the individuals exposed to MRSA cases in hospitals to minimize loss of quality-adjusted life years. We solve the model optimally using data from clinical literature and conduct sensitivity analyses on key parameters, including disutility values and disease parameters, such as prevalence and transmission probability. The optimal screening decisions recommends varying both the frequency and timing of screenings based on initial screening results and room configurations. Particularly in cohort rooms, where the risk of transmission during isolation is increased, initiating screening early is crucial. Although implementing these optimal screening decisions may present challenges due to their complexity, they offer valuable insights for improving MRSA management in healthcare settings, potentially leading to better health outcomes and cost savings compared to current practices. We evaluate and compare the performances of several surveillance policies, including different testing schedules and modalities using the proposed modeling framework enabling us to incorporate the test sensitivities on different days. We suggest alternative screening policies that balance cost-effectiveness with clinical efficacy, aligned with decision-makers' objectives. Furthermore, the model's applicability extends beyond MRSA to other hospital-acquired infections with similar surveillance screening protocols, demonstrating its potential for infection prevention and control strategies.

4 - Political Affiliation and Mental Health During Crises: Insights from Online Communication

Saeede Eftekhari, Tulane University, New Orleans, LA, United States, Zhasmina Tacheva, Claire Senot

The world is struggling with multiple crises on a global scale: from the recent global COVID-19 pandemic and a global recession to the uncertainty of the AI revolution, multiple zones of active conflict, and growing political division. Using advanced analytical techniques, including longitudinal econometric modeling, transformer-based language model analysis, and topic modeling, this paper studies whether political affiliation plays a role in the level of mental health well-being in the period between 2019 and 2023, as manifested in the online communication of the United States-based Twitter users. We found evidence for this differential effect and showed that it depends on multiple factors, including the state in which an online user resides and the type of COVID-related policies in effect in their place. This research contributes to the emergent management science domain of building healthier information ecosystems.

WC46

Summit - 438

Data-driven Analytics and Decision-Making in Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Jaeseung Baek, Northern Michigan University, Marquette, MI, United States

Co-Chair: Mehmet Turkoz, William Paterson University, New Brunswick, NJ, United States

1 - Data-Driven Method for Quality Monitoring in Multi-Mode Surface Topography

Jaeseung Baek, Northern Michigan University, Marquette, MI, United States, Myong Kee Jeong, Elsayed Elsayed

Recent advancements in optical measuring instruments have led to an increased utilization of surface topographic data for quality monitoring of engineered surfaces. This study introduces efficient anomaly detection methods for monitoring variations in different types of surface topography. We explore approaches for monitoring two types of surfaces: single mode and multimode surface topography. Initially, we present a residual-based surface segmentation approach to identify local surface changes in surfaces under a single in-control mode. Subsequently, we introduce a multimode surface prediction model, characterizing the typical behavior of normal surfaces with multiple in-control modes.

2 - Hlfgp-Mobo: Heterogeneous Latent Factor Gaussian Process for Multi-Objective Bayesian Optimization of Longitudinal Data

Aditya Satish Rane, Oklahoma State University, Stillwater, OK, United States, Md Shihab Shakur, Srikanthan Ramesh, Akash Deep

Fabricating functional bioprinted constructs is a time-consuming and expensive process, often involving repeated measure designs. In these designs, post-printing cellular functions, such as albumin and urea secretion, are measured temporally. The goal is to develop a design strategy that effectively models and optimizes multiple outcomes simultaneously. Multi-objective Bayesian optimization (MOBO) typically uses independent Gaussian processes as surrogate models, which assume an independent covariance structure between objectives. This difficulty in capturing correlations can lead to suboptimal results when the objectives are correlated. Additionally, classical methods for modeling panel data such as Generalized Linear Mixed-Effects Models, Structural Equation Modeling, and Latent Growth Models have limitations. These include capturing non-linear effects, computational challenges with larger sample sizes, and addressing measurement errors due to sub-group heterogeneity. To overcome the challenges associated with modeling and optimizing longitudinal data with multiple objectives, we propose HLFGP-MOBO, a hybrid integration of multi-output Gaussian process covariance structure and MOBO. Our methodology comprises of three key components: (1) Model correlated time-varying multiple outcomes by capturing auto-covariance and cross-covariance. (2) Incorporating cellular endpoint sampling noise to adapt heterogeneous noise patterns, accounting for different levels of noise variability. (3) Formulating an adaptive experimentation framework with MOBO acquisition function to iteratively design experiments based on MOGP estimates. We demonstrate the effectiveness of our approach through extensive experimentation in 3D bioprinting, where time-dependent cellular endpoints are collected to measure cell viability and functionality. By optimizing bioprinting process parameters using our methodology, we aim to enhance cell proliferation.

3 - Sequential Partial Least Squares Regression for Multi-Variate Time-Series Sensor Data.

GyeongTaek Lee, Gachon University, SeongNam, Korea, Republic of

In manufacturing environments including semiconductor manufacturing, from the equipment, a variety of sensor data are collected. Process engineers manage manufacturing quality using predictive models based on the collected data . It has been a challenging task to handle this temporally characteristic data, as it is uncertain when patterns of defects might occur. Over the past decade, deep learning has advanced, allowing for the utilization of models that can incorporate temporal characteristics in manufacturing process. However, in manufacturing process. issues such as process drift, maintenance, and changes in process techniques cause periodic changes in data characteristics, making it difficult to secure large amounts of training data. This study proposes a sequential partial Least Squares Regression that can reflect temporal characteristics to enhance the performance of defect detection models. The proposed methodology exhibited higher performance than conventional machine learning and deep learning models in situations with limited data.

4 - Fast Bayesian Support Vector Data Description

Support Vector Data Description (SVDD) is a support vector-based learning algorithm that is used to detect anomalies. It obtains a spherically shaped boundary around the target data and is a very attractive kernel method as it outperforms comparable methods in problems that involve the detection of anomalies. However, training an SVDD algorithm is computationally expensive, since it requires solving a quadratic programming problem with time complexity of order $O(N^3)$. The same problem exists for Bayesian SVDD (BSVDD) processes. In BSVDD applications that involve big and high-dimensional data, a significant disadvantage is the high computational cost since the whole training data set must be stored and evaluated. Also, speed during testing is important in a growing number of applications (e.g., edge computing). To address this problem, we propose a new Fast BSVDD procedure which significantly improves the computational time of traditional BSVDD.

WC47

Summit - 439

Healthcare Systems: Mechanism Design and Operations Management

Invited Session

Health Applications Society

Chair: Zhan Pang, Purdue University, West Lafayette, IN, 47907, United States

Co-Chair: Linggang Qi, Soochow University, Suzhou, N/A

Co-Chair: Jia Zhao, University of Chinese Academy of Sciences, Beijing, N/A

1 - A Geometrically Convergent Solution to Spatial Hypercube Queueing Models for Emergency Services

Yixing Wen, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

The spatial hypercube queueing model was developed 50 years ago to solve spatial queueing problems and has been widely applied to emergency services such as police and ambulance systems. We devise an alternative solution to the hypercube model, and show that it yields the same solution as the original model. We prove our algorithm converges at a geometric rate to an exact solution. While the original model was developed for homogeneous service rates, we extend our method to accommodate heterogeneous service rates and prove the conditions for convergence. Through comprehensive numerical experiments using two datasets from St. Paul, MN, and Greenville County, SC, we show that our approach is more than 80% faster than the original approach with an average percentage error of less than 0.1%. Additionally, we develop a parallel computing algorithm leveraging the convergence property of our approach and two key structural properties of the problem to further reduce its computational time, achieving more than 94% of parallelization.

2 - Wasserstein Risk-Sensitive Appointment Scheduling under Delay Constraints

Jia Zhao, University of Chinese Academy of Science, Beijing, China, People's Republic of

We consider a risk-averse appointment scheduling problem for a healthcare service delivery system with discrete appointment times and service-duration ambiguity, focusing on the tradeoff between overtime work and patient delays. We employ conditional value-at-risk (CVaR) to measure the risk, and to address the ambiguity, we develop a distributionally robust optimization (DRO) model over a type-1 Wasserstein ball centered at an empirical service-duration distribution. We show that the Wasserstein DRO model enjoys a regularized reformulation, which captures an empirical-risk effect and a position-differentiated effect on synthesized aversion to risk and ambiguity. The regularized reformulation provides an important operational insight of position-differentiated robustness for delay management under service-duration ambiguity, without incurring additional complexity. Leveraging the regularized reformulations, the optimal schedule can be determined efficiently by evaluating the empirical CVaRs with binary search for each position sequentially. We then perform sensitivity analysis, primal-dual analysis, and derive finite-sample and asymptotic performance guarantees for the optimal schedule obtained. Furthermore, we extend our model to incorporate sequence decisions for multiple types of patients, which also exploits the regularization structure and can be solved as one instance of mixed-integer linear program. Finally, numerical experiments demonstrate the insights and performance of the proposed approach.

3 - Personalized Health Check-up Decisions Using Data Analytics and Markov Decision Processes

Tianning Zhu, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Guohua Wan

Annual health check-ups are common in many countries but they may not be necessary in case of persons with good health conditions. Using a large data set of health check-ups in China, we analyze the relationships between various indices of the items in health check-ups. With these relationships we develop a partially observable Markov decision process for personalized health check-up decisions so as to maximize the person's expected quality-adjusted life years.

4 - Personalized Lung Cancer Screening Strategies: A Markov Model Approach Using Real Patient Data from China

Ting Wu, Nanjing University, Nanjing, China, People's Republic of

Lung cancer ranks first in incidence and mortality among malignant tumors both in China and globally, often being detected at an advanced stage with a five-year survival rate of 17.4%. Early definitive surgery remains the most effective treatment, highlighting the importance of early detection and screening before symptoms arise and curative treatment is possible. This study explores how to personalize lung cancer screening policies to maximize long-term patient benefits. We first establish a Markov model, utilizing real patient data to estimate parameters and simulate the outcomes of lung cancer screening and subsequent treatment. We then integrate various data, supplement different actions and their benefits, and input this information into a Partially Observable Markov Decision Process (POMDP) model to determine the optimal screening strategy.

5 - Wasserstein Risk-Sensitive Appointment Scheduling under Delay Constraints

We consider a risk-averse appointment scheduling problem for a healthcare service delivery system with discrete appointment times and service-duration ambiguity, focusing on the tradeoff between overtime work and patient delays. We employ conditional value-at-risk (CVaR) to measure the risk, and to address the ambiguity, we develop a distributionally robust optimization (DRO) model over a type-1 Wasserstein ball centered at an empirical service-duration distribution. We show that the Wasserstein DRO model enjoys a regularized reformulation, which captures an empirical-risk effect and a position-differentiated effect on synthesized aversion to risk and ambiguity. The regularized reformulation provides an important operational insight of position-differentiated robustness for delay management under service-duration ambiguity, without incurring additional complexity. Leveraging the regularized reformulations, the optimal schedule can be determined efficiently by evaluating the empirical CVaRs with binary search for each position sequentially. We then perform sensitivity analysis, primal-dual analysis, and derive finite-sample and asymptotic performance guarantees for the optimal schedule obtained. Furthermore, we extend our model to incorporate sequence decisions for multiple types of patients, which also exploits the regularization structure and can be solved as one instance of mixed-integer linear program. Finally, numerical experiments demonstrate the insights and performance of the proposed approach.

WC48

Summit - 440

E-commerce Order Fulfillment: Matching, Inventory Replenishment and Demand Spillover

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Linwei Xin, Booth School of Business, University of Chicago, Chicago, IL, United States

1 - Balancing Demand Spillover and Lost-Sales Under Fulfillment Flexibility

Jingyuan Yang, Booth School of Business, University of Chicago, Chicago, IL, United States, Linwei Xin, Yuan Zhong

One key aspect of inventory management in online retailing is demand spillover, where a stockout at one fulfillment center (FC) leads to demand shifting to another FC. Our research challenges the widely accepted benefits of inventory pooling as described by the square-root law, a concept originally presented in the influential paper by Eppen (1979). This law advocates for efficiency gains through the reduction of system-level safety-stock by pooling inventory across multiple locations, thereby changing its order of magnitude from N to √N, where N represents the number of FCs. We demonstrate that pooling can lead to significant spillovers, with the magnitude growing linearly with N. We further investigate the trade-off between minimizing lost-sales and reducing spillover in two network designs with limited flexibility, and challenge the common wisdom regarding the effectiveness of long chain, a concept related to the design of flexible processes as proposed in the seminal paper by Jordan and Graves (1995). We argue that adopting the design of long chain can significantly increase spillover. In fact, we prove that as we approach the limit of minimizing lost-sales, each unit of reduced lost-sales can lead to as much as N − 1 units of extra spillover. Our study reevaluates both these established principles due to the implications of spillover in modern online retail contexts, and our results shed light on insights for e-retailers to strike an effective balance between spillover and lost-sales especially given the significant cost implications of spillover.

2 - Hybrid Fulfillment Network Operations: Inventory Management and Fulfillment

Nayeon Kim, Wayfair, Boston, MA, United States, Jason Chen

In e-commerce operations, oftentimes an online retailer operates its own fulfillment network as a service for suppliers or to fulfill its own inventory. Fulfillment-by-Amazon and Wayfair-Castlegate are examples of such operations. Some suppliers using the service also run dropship operations from their own or 3PL fulfillment network to complement the online retailer's fulfillment operations. Such hybrid fulfillment networks can offer benefits such as enhanced demand coverage without significant asset investment.

In the context of hybrid fulfillment networks, the optimal inventory replenishment, positioning, and fulfillment decisions require full information and control of both networks, which is often not feasible. For example, the online retailer may not have visibility into the supplier's on-hand or on-order information, or the supplier may only want to leverage a portion of the online retailer's fulfillment network. We will present different inventory replenishment and positioning strategies for hybrid fulfillment networks under different information and control levels and evaluate the pros and cons of each.

3 - A Deep-Learning Approach to High-Dimensional Impulse Control: with Applications to Inventory Management

Wouter van Eekelen, University of Chicago, Booth School of Business, Chicago, IL, United States, Baris Ata, Yuan Zhong

We consider a class of joint replenishment problems with renewal demands, in which procurement fixed costs can be saved by replenishing a group of different types of items together. So far, solving these problems numerically for a large number of items has been hindered by the dreaded "curse of dimensionality." To tackle this challenge, we introduce an approximation method via impulse control---a stochastic control problem that allows the system controller to intervene through discrete jumps in the state space. We then propose a novel deep-learning framework in order to solve this impulse control approximation to the original inventory problem. Grounded in the theory of backward stochastic differential equations (BSDEs), our framework relies crucially on probabilistic identities that highlight a deep connection between impulse control and stochastic target problems.

4 - Boosting Deep Reinforcement Learning in Practice: The "Less is More Approach"

Linwei Xin, Booth School of Business, University of Chicago, Chicago, IL, United States, Yaqi Xie, Will Ma

There is a prevailing perception that black-box deep reinforcement learning (DRL) algorithms can effectively solve OR/OM problems. However, in this talk, we present preliminary evidence that challenges this perception. We demonstrate that leveraging problem-specific structures can significantly boost the performance of DRL algorithms in the context of inventory management. Our key message is that combining black-box DRL with problem-specific structures not only improves outcomes but also highlights the tangible value of moving beyond purely black-box approaches.

WC49

Summit - 441

Queueing Applications for Service Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Tugce Isik, Clemson University, Clemson, SC, United States

1 - Optimal Admission Control in Queues with Multiple Customer Classes and Abandonments

Runhua Wu, ISyE Georgia Tech, Atlanta, GA, United States, Hayriye Ayhan

We consider a Markovian, finite capacity queueing system with multiple customer classes and multiple servers where customers waiting in line can get impatient and leave without being served. There is a cost associated with abandonments and a holding cost associated with customers in the system. Admitted customers pay a class dependent reward at the time of arrival. Under these assumptions, our objective is to characterize the optimal admission control policy that maximizes the long-run average reward. We formulate the problem as a Markov decision process problem and prove that the optimal policy is a double threshold policy. We also identify sufficient conditions under which the optimal policy reduces to a threshold policy which is known to be optimal for systems without abandonments. After investigating how the optimal long-run average reward changes with respect to system parameters, we conclude with a comparison of the performance of the optimal threshold policy to the optimal policy.

2 - Stochastic Systems Fed by Self-Exciting Point Processes

Brian Fralix, Clemson University, Clemson, SC, United States, Jeffrey Kharoufeh, David Pittman

It is well-known that the infinite-server queue fed by a nonhomogeneous Poisson process is remarkably tractable and can be used to establish good staffing levels for multiserver queueing systems. In this talk, we present a collection of self-exciting point processes, where both infinite-server queues and shot-noise models fed by these processes possess various distributions and performance measures that are surprisingly tractable.

3 - Optimal Control in Two-Class Bass Diffusion Mode

Hojun Choi, Colorado School of Mines, Golden, CO, United States, Ohad Perry, Achal Bassamboo

We propose and consider a control problem for a clearing system in the setting of healthcare with two population classes. Each class consists of two states: susceptible and infected. Furthermore, classes transition into infected state either spontaneously (e.g., interaction with the outside system) or through interactions with the infected population. We formulate the problem as a multi-class queue-clearing model. The control is the instantaneous allocation of scarce resources such as vaccine to each class at every point in time. The objective is to find a control policy that minimizes the cumulative amount of the infected when the system clears. We first show that an optimal policy has the Bang-Bang structure, i.e., we should allocate the resources to one class completely at a given point in time. However, the explicit solution is intractable due to the nonlinear dynamics of the system. Therefore, we propose a strict priority policy and a practical policy. A strict priority policy allocates all the resources to one class until it clears before switching, while a practical policy allocates proportional to the remaining susceptible population. We numerically demonstrate the optimality gap for these two controls under reasonable parameters. Our results indicate that a strict priority policy is optimal, and a practical policy can be a compromise with a heavy penalty.

4 - Optimal Routing in Loss Systems with Flexible Servers and Customers

Tugce Isik, Clemson University, Clemson, SC, United States, Avnish Malde

We study a loss system with two classes of servers and a number of customer types. We assume that there is at least one type of flexible customers that can receive service from servers of either class. Arrival rates are dependent on customer type but service rate is identical for all customer types and classes of servers. We formulate the problem as a Markov Decision Process (MDP) with rewards that are dependent on the customer type. For small systems, we use our formulation to characterize the optimal policies. For larger systems, we study the structure of the optimal policies. Our analysis shows that the optimal policies are of threshold type. Using this policy structure, we develop heuristic routing policies for general systems and benchmark these heuristics by greedy policies that route flexible customers to servers of a particular class whenever possible. Our numerical results show that our heuristics perform well across a variety of parameter settings.

WC50

Summit - 442

Advances in Machine Learning and Optimization

Contributed Session

Chair: Ahtesham Bakht, USF, 14306 Wedgewood Circle, Tampa, FL, 33613, United States

1 - Identification and classification of the generated adversarial samples using deep learning and machine learning methods

Ahtesham Bakht, University of South Florida, Tampa, FL, United States, Ankit Shah, Nathaniel Bastian

Intrusion Detection Systems (IDS) powered by artificial intelligence have led to enhanced effectiveness in automatically detecting cyber-attacks. However, the learning models may become cyber-attack targets through carefully crafted adversarial samples (also called evasion attacks). Recent advancements indicate that deep learning models are susceptible to adversarial samples. These samples are carefully crafted by introducing slight alterations to normal samples, which are undetectable to humans but can potentially deceive the deep learning models, leading to erroneous or desired outputs by the adversary. Such attacks enable adversaries to circumvent the IDS, leading to delayed detection of attacks and infrastructure damage.

In this study, we use five state-of-the-art adversarial generation methods such as Deep PackGen, autoencoder-based generative model, Projected Gradient Descent (PGD), Fast gradient sign method (FGSM), and Deep-Fool, on the CICIDs- 2017 dataset and after creating the adversarial samples we applied the five mainstream stream machine learning and deep learning methods to identify and classify each of the adversarial generated samples. Some of the adversarial sample generation is novel in the sense that it does not change the functionality of the packets.

The findings outlined in this paper offer valuable perspectives for designing systems geared toward identifying adversarial samples. In the larger sense, it would help us identify the source of evasion attacks. Additionally, they can serve as a blueprint for devising novel approaches to detect such samples.

2 - Sample-Efficient Agnostic Boosting

Karan Singh, Tepper School of Business, Pittsburgh, PA, United States, Udaya Ghai

The theory of boosting provides a computational framework for aggregating approximate weak learning algorithms, whose performance is marginally better than that of a random predictor, into an accurate strong learner. In the realizable case, the success of the boosting approach is underscored by a remarkable fact that the resultant sample complexity matches that of a computationally demanding alternative, namely Empirical Risk Minimization (ERM). This in particular implies that the realizable boosting methodology has the potential to offer computational relief, without compromising on sample efficiency.

Despite recent progress, in agnostic boosting, where joint distributional assumptions on the labels and feature descriptions are absent, ERM outstrips the agnostic boosting methodology in being quadratically more sample efficient than the all known agnostic boosting algorithms. In this paper, we make progress on closing this gap, and give a substantially more sample efficient agnostic boosting algorithm than those known, without compromising on the computational (or oracle) complexity. A key feature of our algorithm is that it leverages the ability to reuse samples across multiple rounds of boosting, while guaranteeing a generalization error strictly better than those obtained by blackbox applications of uniform convergence arguments. We also detail resultant implications for previously studied learning problems.

3 - Single-Attempt Hyperparameter Optimization Using Graph Neural Networks

Chanho Kim, Sungkyunkwan University, Suwon, Korea, Republic of, Dohyun Bu, Bonyoung Koo, Yulim So, Gaeun Kim, Woo-Jae Lee, Jong-Seok Lee

As deep learning has gained significant attention, selecting the best hyperparameter configuration for neural network models has become crucial for optimal performance. Consequently, hyperparameter optimization (HPO) has emerged as an important problem. Many recent advancements in methodologies and theories have been developed to address HPO. However, despite efforts to reduce the search space and find efficient paths to the optimum, they still require substantial time and computational resources due to trial-and-error processes. In this research, we approach the HPO problem as a recommendation problem, assuming that optimal hyperparameter settings are similar across similar tasks. Our goal is to achieve near-optimal settings in a single attempt. Our method combines graph neural networks to represent the relationships between tasks and hyperparameter settings, recommending the best configuration. Additionally, Bayesian optimization can be applied to further improve the recommendations and systematically update the task-hyperparameter history if necessary. Numerical experiments show that our method identifies individual optimal settings with fewer trials and infers near-optimal settings for unseen tasks in a single attempt.

4 - First-improvement or best-improvement? An in-depth local search computational study to elucidate a dominance claim

Celso Ribeiro, Universidade Federal Fluminense, Rio de Janeiro, Brazil, Robin Moine, Daniel Aloise, Jonathan Jalbert

Local search methods start from a feasible solution and improve it by successive minor modifications until a solution that cannot be further improved is encountered. They are a common component of most metaheuristics. Two fundamental local search strategies exist: first-improvement and best-improvement. In this work, we perform an in-depth computational study using consistent performance metrics and rigorous statistical tests on several classes of test problems considering different initialization strategies, neighborhood structures, and types of tested instances to evaluate whether one strategy is dominant over the other. The numerical results show that computational experiments previously reported in the literature claiming the dominance of one strategy over the other for the TSP may not be consistent, and their extrapolation to other problems is not correct, possibly due to the use of a less rigorous experimental setting. Still, our results reinforce the need for extensive experimentation to decide the most appropriate strategy for each specific problem and context.

5 - Structured Real-time bidding algorithm with ESS control: Cooperative Multiagent offline RL Approach

Wooje Seong, Kia, Seoul, Korea, Republic of, Euncheol kim, Wonjong Lee, Yoonmo Koo

As renewable energy deployments accelerate, so does the need for a Virtual Power Plant (VPP) that optimizes the operation of distributed energy resources. This study proposes a Reinforcement Learning (RL)-based bidding and operation model that maximizes the profits of VPPs given uncertain power market operations. In order to optimize resources in a structured market framework—where profit is determined both by bidding and operational decisions—this study adopts a bi-level, multi-agent RL approach and the off-line learning structure. Specifically, the model considers the sequential and cooperative decision-making of bidding and operations by reflecting the communication layer between each agent and implicitly assigning the credit of individual bidding and operational decisions to the final total reward. By implementing this Centralized Training but Decentralized Execution (CTDE) architecture, we effectively derive the cooperative decision-making of the two agents to optimize the total revenue of the VPP rather than maximize the rewards of distinct bidding and operational decisions. Moreover, in order to further exploit the practicality of the proposed model, we constructed an offline-RL structure that utilizes the expert decision data derived from the history of complete market information for training. Lastly, we conducted a case study for South Korean power market and the results produced bidding and operational revenue streams that were nearly identical to revenues derived from expert decisions made with the benefit of perfect information, despite its having been applied to an uncertain market case.

6 - Increasing operations efficiency using LLMs and GPU acceleration

Hugo Linsenmaier, NVIDIA, Santa Clara, CA, United States, Victor Miara, Alex Fender

Optimizing supply chain operations is a time-consuming process requiring deep technical expertise in mathematical modeling. Understanding the outcome of the optimized plan also relies heavily on human expertise. This process can be very costly as plans need to be updated frequently to account for disruptions. In this work, we propose an LLM agent for mathematical models that accelerates operations by transforming natural language operations queries into actionable optimized strategies. Our agent can fetch relevant historical data, solve the formulated problem using GPU solvers, and answer complex queries on potential business outcomes. The efficiency of our stack makes it possible to evaluate different scenarios in near real time on large scale problems.

WC51

Summit - 443

Advancements in Medical Decision Support Systems

Contributed Session

Chair: Helena Müller, University of Cologne, Cologne, Germany

1 - Towards Explainable Patients Recovery Prediction in Neuro Critical Care Unit

Eleanor Belkin, Johns Hopkins University Whiting School of Engineering, Baltimore , MD, United States, Dr. Tamas Budavari, Dr. Rohan Mathur, Lin Cheng, Peter Dziedzic, Niteesh Potu, Eusebia Calvillo, Josiah Lim, Vishank Shah, Jose Suarez

Our work aims to improve triage and care for patients who underwent elective craniotomies for brain tumor resection. These patients are typically admitted to the Neurosciences Critical Care Unit (NCCU) post-operatively. However, only 15-20% of these patients develop conditions in the post-operative setting that require interventions that can only be performed in the NCCU. Given the scarcity of NCCU beds and the high costs associated with critical care, there is a need to optimize these resources. Our work focuses on statistically explainable machine learning (ML) models. We explore Bayes classifiers, boosted decision trees, logistic regression, and Generalized Additive Models (GAMs) to establish a baseline. We use pre-operative EMR data and brain scans to predict the need for NCCU care post-operatively. We integrate statistical classification methods with the optimal features extracted by the deep learning network from patients' brain scans, enabling more accurate predictions while maintaining interpretability of the model. Our approach addresses the limitations of both methods, offering a hybrid system with the advantages of each and allowing to develop a novel tool for the classification by utilizing a probabilistic outcome of the model.

2 - Posturography-Based Decision Support System for mTBI Detection

Eren Darici, Western Michigan University, Kalamazoo, MI, United States, Alessander Danna-dos-Santos, ILGIN ACAR, Kira Hamelink

A mild traumatic brain injury (mTBI) or concussion, as defined by the CDC, is the disruption of normal brain functions that can be caused by a bump, blow, or jolt to the head. This study leverages custom-made force plates to collect postural sway data from over 120 patients to create a decision support-system-like approach to the diagnosis of mTBIs using various classification models. Despite accounting for roughly 90% of all traumatic brain injuries, not all mTBIs are diagnosed or treated. If left untreated, mTBIs can lead to post-concussion syndrome and a number of long-term cognitive impairments. Another long-term effect of mTBIs is the inability to adequately control postural sway; trauma to brain centers responsible for maintaining balance and proprioception reduce the ability to stabilize when standing upright. Postural sway can be measured and assessed with posturography. Analysis of posturography results can provide insight into the physical condition of a patient's brain following an mTBI or aid in the diagnosis of previously undetected mTBIs. In a clinical setting, however, it becomes challenging and time-consuming to conduct the structural analysis of posturography manually. With a high accuracy of 88%, this study examines a novel set of methodologies for mTBI detection that are mobile, inexpensive, and less time-consuming than traditional methods. The system resulting from this study has the potential to greatly increase the diagnosis rate of mTBIs, reducing the impact of untreated mTBIs. Future work includes the use of this system for the detection of neurodegenerative diseases, such as Alzheimer's and Parkinsons.

3 - Temporal Framing Effects on Responses to Messages Promoting Social Distancing: The Moderating Role of Involvement

Giha Shin, Yonsei University, Seoul, Korea, Republic of, Jarim Kim, Jihyun Lee

This study investigated how individuals' level of involvement moderates temporal framing effects in the context of COVID-19 social distancing, which provides a unique context in which most people have high involvement. Eight hundred participants were randomly assigned to one of two conditions (temporal framing: proximal vs. distal). Participants' involvement was operationalized as high or low based on whether they had close relationships with someone with an underlying condition. When exposed to distally framed messages, participants with high involvement reported more favorable attitudes and greater intentions toward social distancing than individuals with low involvement. The effects were mediated by response efficacy, concern for others, and anticipated regret. When exposed to proximally framed messages, participants with different levels of involvement did not report different responses. The study makes theoretical contributions to health care literature and practical contributions to designing effective health communication messages.

4 - The Impact of Consecutive Working Days on Individual Nurse Fatigue: An Analysis on Neonatal Wards

Helena Müller, University of Cologne, Cologne, Germany, Michael Becker-Peth, Ludwig Kuntz

Purpose: Previous research has shown the negative impact of fatigue on different outcomes in hospital settings. Our analysis contributes to this research with a longitudinal dataset at the individual nurse level, showing empirically that the operational metric of consecutive working days is correlated with the perceived fatigue.

Methodology: The data were collected as part of the Neo-CamCare study using a combined approach of an observational study and a standardised diary questionnaire. A total of 1,191 daily questionnaires from 65 nurses across three neonatal wards were collected over six-month periods on each ward. Panel regression models with fixed effects were used for analysis. To test for endogeneity, an instrumental variable approach was employed.

Results: Initial results of the survey data indicated that a higher number of consecutive working days without breaks was significantly associated with higher individual fatigue levels. Further moderation analyses revealed that the effect of consecutive working days on fatigue was contingent upon the specific shift type.

Implications: The results showed empirically that the number of consecutive working days can be used as an objective indicator of fatigue among nurses. Thus, the analysis has implications for nurse scheduling from a managerial perspective as limiting the number of consecutive working days could reduce the perceived fatigue level. Our research highlights the importance of understanding the dynamics between fatigue and work schedules among nurses, providing valuable insights for hospital management aimed at optimizing nurse scheduling while considering the perceptions of nursing staff.

WC52

Summit - 444

Advances in Operations Management

Invited Session

MSOM: Supply Chain

Chair: Zihao Qu, The University of Massachusetts Amherst, Belchertown, MA, United States

1 - Probabilistic Design in the Operation of Service Systems

Jiangze Han, University of British Columbia, Vancouver, BC, Canada, Christopher Ryan, Xin T. Tong

Motivated by the design and operation of service systems involving non-parametric probability distributions, we study optimization problems in the space of probability measures. We propose an implementable gradient descent method and tailor it to two applications: user profile decomposition, which is an offline constrained optimization problem, and matching spatial supply with spatial demand, an online optimization problem. We analyze the performance of our method in both settings, demonstrating that (i) it converges to a solution satisfying the proposed optimality condition in the offline application, and (ii) achieves sublinear static and dynamic regret in the online setting.

2 - Sound of Silence: when to Conceal Attribute Information? the Roles of Consumer Search, Inventory, and Channel Structure

Xingyu Fu, University of New South Wales, Sydney, Australia, Zihao Chen, Pin Gao, Ying-Ju Chen

This study examines a seller's joint information disclosure and pricing decisions when launching a new product with an easy-to-communicate objective attribute and a hard-to-describe subjective attribute. We investigate how consumer search cost, inventory level, and channel structure influence the seller's optimal information policy. For brick-and-mortar retailing, our analysis suggests that the seller should conceal (or disclose) information when the search cost is low (or high). Interestingly, when the inventory level is high, the seller conceals information to retain the diamond-in-the-rough consumers, aiming to match the abundant inventory with high demand. Conversely, when inventory is low, the seller discloses information to dissuade some consumers with low objective valuations from searching, thereby reducing demand and alleviating concerns about product unavailability. In other words, the seller uses information to balance search and stock. The structure of optimal information policy is robust after considering several extensions such as partial disclosure, efficient allocation, and atomic consumers. In addition, surprisingly, fiercer demand-side competition (e.g., larger consumer population) can reduce the seller's revenue due to potential price markdowns to ease consumer concerns about competition. For omnichannel retailing with an additional online purchasing channel, our analysis suggests that abundant inventory can lead to a contrary information policy from the offline business. Collectively, our study provides managerial insights into how demand-side factors (e.g., search cost) and supply-side factors (e.g., inventory and channel structure) impact the seller's optimal information and pricing strategy. It also reveals potential risks associated with intensified demand-side competition.

3 - Dynamic Pricing and Resource Allocation in a Multi-Server Queue

Zerui Wu, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Xu Sun, Ran LIU

We explore dynamic pricing and resource allocation policies for a multi-server queue serving heterogeneous customer classes with varying price and delay sensitivities, and class-dependent service rates. This setting is relevant to applications such as drone delivery, EV charging stations, and cloud computing, where timely response and service are essential. We consider a manager who makes real-time pricing and resource allocation decisions to maximize long-run average profit. Due to the high-dimensional system state, finding an optimal solution is intractable. We propose a tractable myopic policy that solves a single-step optimization problem at each decision epoch. We prove that this simple policy is asymptotically optimal, with an O(n^(-2/3) log n) optimality gap in both revenue and delay costs compared to the optimal fluid solution, where n is the number of servers. Notably, the policy may exhibit non-monotonic pricing with respect to congestion when queues are observable to customers. Our research provides an easy-to-implement approach for real-time operations of many-server queues while maintaining near-optimal performance.

4 - Decentralized Insurance

Baikun Leng, University of Washington Foster School of Business, Seattle, WA, United States, Rowena Gan

In the traditional insurance industry, the evaluation of claims is often biased due to the inherent interest of claim assessors in maximizing company profits. We explore the potential of decentralized insurance, powered by blockchain technology, as an alternative approach to mitigate such biases. In decentralized insurance models, claims are evaluated by a "court" of assessors who are independent of the insurance company. These assessors vote collectively to determine the outcome of a claim, with their compensation tied to their alignment with the majority decision. This mechanism incentives assessors to provide judgments that reflect the consensus of the group, rather than the financial interests of the insurance company.

5 - Flexible vs. Dedicated Technology Choice in the Presence of Multi-product Subscription Programs

Liling LU, Singapore Management University, singapore, Singapore, ONUR BOYABATLI, Sarah Yini Gao

This paper studies flexible versus dedicated technology choice and capacity investment decision of a two-product firm under demand uncertainty in the presence of subscription programs. The key feature of subscription programs is that a proportion of customers that are allocated a particular product later switch to using the other product (if available). We build a two-stage stochastic program to study the optimal decisions, and investigate how the correlation between the two subscription demands affect the profitability with each technology, and shape the optimal technology choice decision.

WC53

Summit - 445

Environmental Sustainability and Social Responsibility in Supply Chain

Invited Session

MSOM: Sustainable Operations

Chair: Mili Mehrotra, University Of Illinois Urbana Champaign, Champaign, IL, United States

Co-Chair: Xueze Song, The University of Alabama, TUSCALOOSA, United States

1 - Does the Focus on Workplace Safety Affect Environmental Performance? Analyzing the Impact of Osha Inspections on Toxic Releases of Manufacturing Facilities

Suresh Muthulingam, The Pennsylvania State University, University Park, PA, United States, In Joon Noh

Manufacturing plays a critical role in the U.S. economy but ensuring workplace safety has been a serious concern within the sector. The Occupational Safety and Health Administration (OSHA) was created to promote safe work practices in manufacturing. Since its inception, OSHA has extensively used facility inspections to assess whether manufacturing operations adhere to safety standards. A large body of work has investigated the implications of OSHA inspections from economic and safety perspectives, but hardly any research explores the implications of OSHA inspections from an environmental perspective. We seek to bridge this gap in the literature by examining whether OSHA inspections affect the toxic releases of manufacturing facilities. We draw on the attention-based view of the firm and hypothesize that OSHA inspections will draw managerial attention towards safety-related issues, but at the same time end up diverting attention away from environmental issues. We created a dataset that spans from 1987 to 2016 and includes details on the toxic chemical releases of 33,838 manufacturing facilities across the U.S. Our dataset also includes information on the 36,583 OSHA inspections conducted on these manufacturing facilities. Our econometric analyses show that OSHA inspections lead to increased toxic chemical releases at manufacturing facilities, and that this effect is largely driven by the inspections that cite violations of safety regulations. We additionally find that facility managers undertake fewer modifications to reduce toxic chemical releases at the source when OSHA inspections identify violations, which clarifies the mechanism that drive the increases in toxic releases.

2 - Pooling Carbon Targets

Christian Blanco, Ohio State University, Columbus, OH, United States, Oznur Ozdemir, Erdinc Akyildirim

Many corporations now set voluntary direct (Scope 1) and indirect (Scope 2) carbon emissions reduction targets, but low success rates reflect that firms are not experienced in designing these targets. Does pooling Scope 1 and 2 targets increase the likelihood of success? We use ten years of data collected by the CDP (formerly the Carbon Disclosure Project) to explore our research question.

3 - Fairness in Crowdwork: Making the Human Ai Supply Chain More Humane

Martin Gonzalez Cabello, University of California, Los Angeles, Los Angeles, CA, United States, Charles Corbett, Auyon Siddiq

Digital supply chains, including those that support artificial intelligence (AI), share notable similarities with physical supply chains – including multiple levels of outsourcing, reliance on low-wage labor, and limited transparency – raising ethical concerns about upstream labor conditions. Through a survey of workers on a major digital labor (“crowdwork”) platform, this article aims to shed light on the factors that shape workers’ perceptions of fair treatment and examines how platform design can influence worker welfare. Drawing from physical supply chains, it provides guidance for managers on how to promote responsible labor practices in digital supply chains.

4 - How do Favorable and Unfavorable Audit Outcomes Impact Supplier CSR Risk?

Seb Heese, NC State, Raleigh, NC, United States, Xiaojin Liu, Rob Handfield, Tim Kraft

Imperfections can exist in the design of supplier audits; the audit score a facility receives may not reflect its true compliance ability. This difference can alter the perception of the outcome, and thus, impact a facility's motivation to improve. We examine the influence of outcome favorability, defined as the difference between the audit score a facility received (we refer to this as the reported audit score) and the audit score the facility should have received based on its compliance ability (we refer to this as the corrected audit score), on the facility’s future CSR risk. To empirically address our research question, we use Item Response Theory and a dataset consisting of historical, facility-level audits for a global apparel brand. We develop a measure for CSR risk using the severity score for audit outcomes in the subsequent round of audits. We find that higher outcome favorability helps to reduce future CSR risk. Furthermore, the analysis of our moderating factors shows that this effect is more pronounced in facilities with low compliance ability, facilities that experienced a greater positive change in outcome favorability since their previous audit, facilities that have been audited a fewer number of times, and facilities in developing countries. Buyers are increasingly being pressured to take more collaborative and developmental approaches to improve their suppliers’ CSR. However, these types of efforts can be difficult and are not always successful; we provide guidance to buyers by identifying when suppliers are more likely to be intrinsically motivated to reduce CSR risk.

WC54

Summit - 446

Diversity and Fairness in High-stakes Decisions

Invited Session

MSOM: Service Operations

Chair: Vahideh Manshadi, Yale, Westwood, United States

Co-Chair: Faidra Monachou, Yale University, New Haven, CT, United States

1 - The Feedback Loop of Statistical Discrimination

Jackie Baek, NYU Stern School of Business, New York City, NY, United States, Ali Makhdoumi

We study a dynamic model of interactions between a firm and job applicants to identify mechanisms that can lead to long-term discrimination. In each round, the firm decides which applicants to hire, where the firm’s ability to evaluate applicants is imperfect. Each applicant belongs to a group, and central to our model is the idea that firms become better at evaluating applicants from groups in which they have hired from in the past. We establish the firm's initial evaluation ability to be a critical factor in determining long-term outcomes, and small differences in this quantity across groups can steer them towards drastically different outcomes. Even if two groups of applicants are identical in size and underlying skill distributions, a marginal difference in the firm's initial ability to evaluate applicants from each group can lead to persistent disparities that exacerbate over time through a feedback loop.

Importantly, the dynamic nature of our model allows us to assess the impact of interventions. We show that drastic short-term interventions are preferable compared to milder long-term interventions. Additionally, we find that smaller groups face inherent disadvantages, and hence, enforcing a parity constraint, such as demographic parity, can be ineffective when group sizes differ.

2 - Monoculture in Matching Markets

Kenny Peng, Cornell University, New York, NY, United States

Algorithmic monoculture arises when many decision-makers rely on the same algorithm to evaluate applicants. An emerging body of work investigates possible harms of this kind of homogeneity, but has been limited by the challenge of incorporating market effects in which the preferences and behavior of many applicants and decision-makers jointly interact to determine outcomes.

Addressing this challenge, we introduce a tractable theoretical model of algorithmic monoculture in a two-sided matching market with many participants. We use the model to analyze outcomes under monoculture (when decision-makers all evaluate applicants using a common algorithm) and under polyculture (when decision-makers evaluate applicants independently). All else equal, monoculture (1) selects less-preferred applicants when noise is well-behaved, (2) matches more applicants to their top choice, though individual applicants may be worse off depending on their value to decision-makers and risk tolerance, and (3) is more robust to disparities in the number of applications submitted.

3 - Generative Social Choice

Paul Gölz, Cornell University, Ithaca, NY, United States, Sara Fish, David Parkes, Ariel Procaccia, Gili Rusak, Itai Shapira, Manuel Wüthrich

Traditionally, social choice theory has only been applicable to choices among a few predetermined alternatives but not to more complex decisions such as collectively selecting a textual statement. We introduce *generative social choice*, a framework that combines the mathematical rigor of social choice theory with the capability of large language models to generate text and extrapolate preferences. This framework divides the design of AI-augmented democratic processes into two components: first, proving that the process satisfies rigorous representation guarantees when given access to oracle queries; second, empirically validating that these queries can be approximately implemented using a large language model. We apply this framework to the problem of generating a slate of statements that is representative of opinions expressed as free-form text; specifically, we develop a democratic process with representation guarantees and use this process to represent the opinions of participants in a survey about chatbot personalization. We find that 93 out of 100 participants feel “mostly” or “perfectly” represented by the slate of five statements we extracted.

4 - Diversity in Labor Markets: Implications of the Rooney Rule

Vahideh Manshadi, Yale University, New Haven, CT, United States, Faidra Monachou, Soonbong Lee

In 2003, in response to concerns about the low representation of African Americans in head coaching and management positions, the NFL adopted the Rooney Rule, which requires every team with a vacancy to interview at least two minority candidates. While several organizations, from the tech industry to law firms, have since adopted similar hiring policies, the Rooney Rule has seen limited success in diversifying coaching teams or C-suites. Motivated by the mixed outcomes of the Rooney Rule in practice, we consider a matching model with interviews to investigate the impact of fairness constraints, such as the Rooney Rule, at the interview stage. In our model, candidates belong to two demographic groups, and their match value distributions are symmetric across groups. However, match values are only revealed after the interview. Consequently, the firm makes interview decisions based on observable scores that provide only partial information about the match values. The score distributions are asymmetric across groups, with one group having a higher variance in score distribution. The firm has limited capacity for interviewing and hiring and aims to maximize the total match value within these capacity constraints. We show that imposing constraints like the Rooney Rule at the interview stage does not mitigate the disparity in hiring between the two groups, highlighting the limitations of such interventions. Additionally, we extend our model to a two-firm scenario to study the impact of competition.

WC55

Summit - 447

Relationship between Supply Chain Network Characteristics and its Resilience

Invited Session

Public Sector OR

Chair: Alican Yilmaz, northeastern university, Boston, United States

Co-Chair: Ozlem Ergun, Northeastern University, Newton, MA, United States

1 - Relationship between supply chain network characteristics and its resilience

Alican Yilmaz, Northeastern University, Boston, MA, United States

Supply chain networks involve complex interactions among their constituent entities.These interactions encompass not only the structural properties of the network but also the service-based characteristics of the supply chain. Inspired by this, we have demonstrated that traditional centrality metrics may prove inadequate, especially in scenarios where service characteristics are heterogeneous.Thus, modified flow-based centrality metrics tailored to the specific context can yield more robust results in identifying critical nodes in the supply chain network. In this work, we first developed a linear lost demand model and illustrated that modified flow-based centrality metrics outperform generic centrality metrics in identifying critical nodes. As second part of the work, we focused on the Defender-Attacker-Defender(DAD) models. Unlike attacker-defender(AD) models, DAD models include the defender’s decision of pre-disruption. With that capability, DAD models outperforms AD models in terms of scope and analysis. However, DAD models, owing to their complex structure, often entail long run-times to reach optimal solutions. To address this challenge, numerous exact and heuristic methods have been developed. In our work, we propose that integrating information from these centrality metrics could potentially expedite the run-time of DAD models. Initial findings suggest promising results in this regard.

2 - Supplier's Capacity Decisions and Supply Network Performance under Extreme Uncertainties

Yuhong Li, Old Dominion University, Norfolk, VA, United States, Kedong Chen, Dmitry Ivanov, Anand Nair, Long Xia

Under extreme conditions caused by events like the COVID-19 pandemic, firms require more frequent operational decision-making to adapt to the volatile business environment. Such short-term decision-making is challenging in a fast-changing environment as decision-makers have limited knowledge of potential risks. In this study, we consider the effect of suppliers' *capacity decision*, a critical firm strategy, on supply network performance. We adopt a multi-method approach. First, we conduct text analytics on news articles and demonstrate that a firm's capacity decision during the extreme conditions is positively related to its asset-based performance. Next, using insights from the analysis, we build an agent-based model and design experiments to investigate the appropriate level of capacity decision across suppliers that improves the network performance. We then examine if a focal firm can further improve network performance by nudging suppliers' risk-taking behavior based on the tier that they belong to or their centrality in the supply network. The findings suggest that network performance can further improve when tier-1 suppliers take higher capacities than tier-2 suppliers, and when suppliers with higher degree centrality take higher risk preferences than less central suppliers. These findings provide insights for addressing extreme conditions by managing risk preference.

3 - Mitigating Worst-Case Supply Chain Disruptions via Safety Stock Pre-Allocation

Carson Grose, Northeastern University, Boston, MA, United States

Engineering resilient supply chains is crucial to ensuring critical commodities such as fuel, pharmaceuticals, and emergency response supplies are available when needed. A resilient supply chain should function even in the event of a disruption, which can come in the form of natural disasters, equipment failures, or attacks by nefarious actors. One way to build resilience in a supply chain is through the pre-allocation of inventory, or safety stock to carefully chosen locations in the distribution network. Safety stock is especially effective at mitigating short term supply chain disruptions. Resilience through safety stock comes with a cost however, so it is important to optimally allocate inventory to make the most effective use of a limited budget. This work studies a protection-interdiction-operation framework applied to an n-echelon supply chain model. Our model determines a worst-case network disruption and a corresponding best-case safety stock allocation plan. This enables us to quantify the worst-case cost that a disruption could incur for given safety stock budgets. We use a column-and-constraint generation (C&CG) algorithm to exactly solve the model. Our results reveal managerial insights that quantify the relationship between levels of protection investment and resilience.

4 - Improving supply chain network resilience: A framework for diagnosing vulnerabilities and identifying adaptation strategies

Chelsey Graham, Massachusetts Institute of Technology, Cambridge, MA, United States, Jarrod Goentzel

We investigate the resilience of supply chain networks, concentrating specifically on the robustness of distribution networks and the process of diagnosing the criticality of nodes. The methodology centers on the vulnerability that nodes can induce on the network, using a three-step approach: 1) Identifying a set of Key Performance Indicators (KPIs) for gauging network vulnerability, 2) Measuring vulnerability for randomly generated supply chain networks with various degree distributions via disruption simulation, and 3) Defining adaptation strategies, through re-optimization and rewiring of the network, as well as practical measures a firm could deploy in the event of a disruption. We conclude by highlighting the pivotal role of network structures in creating and preserving network robustness and unveil potential adaptation strategies for businesses working to reinforce their supply chains against unforeseen disruptions.

WC56

Summit - 448

Solutions to Build Climate Resiliency

Contributed Session

Chair: Farzane Ezzati, University of Houston, Houston, United States

1 - Enhancing Energy Resilience in Vulnerable Communities via ES+Solar Microgrids

Farzane Ezzati, University of Houston, Houston, TX, United States, Zhijie Dong, Gino Lim

Climate change is exacerbating weather-related disruptions to electrical grids, aggravating the risk of prolonged power outages that endanger the safety and well-being of all populations, particularly vulnerable communities. While renewable energies have received great attention for their potential to meet community needs during grid failures, their adoption requires efficient investment management while prioritizing the enhancement of energy resilience.

In our study, we propose a community microgrid that combines energy storage (ES) and Solar power to bolster the energy resilience of vulnerable households. To assess the efficacy of the proposed solution, we model the ES+Solar community microgrid as a two-stage stochastic programming model, accounting for various uncertainties including power outage duration, load demand, solar power generation, and future load demand growth. Additionally, we define four energy resilience metrics including Impact Endurance, Sustained Access, Demand Assurance, and Peak Assurance to comprehensively evaluate the energy resilience benefits of the proposed solution from various perspectives. Furthermore, our study advocates for the implementation of Demand Response Programs (DRP) to further support the emergency power supply.

The model is subsequently optimized using cutting-edge optimization solvers. Our numerical analysis serves two primary objectives: firstly, to provide guidance on the installation of the ES+Solar community microgrid and assess its resilience performance; and secondly, to offer valuable informative and prescriptive managerial insights aimed at evaluating energy resilience and various factors influencing the investment decisions.

2 - Cooperation in Transmission Expansion Planning: Enhancing Grid Reliability and Efficiency Under a Changing Climate

Kerem Ziya Akdemir, Pacific Northwest National Laboratory, Richland, WA, United States, Kendall Mongird, Jordan Kern, Konstantinos Oikonomou, Nathalie Voisin, Casey Burleyson, Jennie Rice, Mengqi Zhao, Cameron Bracken, Chris Vernon

Electricity grids are challenged to maintain reliability during more intense and frequent extreme weather events due to climate change. This challenge is exacerbated by multi-sector electrification and power sector decarbonization through increased reliance on variable renewable energy, which necessitates the expansion of transmission infrastructure. However, transmission expansion planning is often complicated by intertwined planning authorities and jurisdictions, and allocation of large capital investment needs. These factors cause authorities to manage transmission investments individually (i.e., only/mostly intraregional planning), which can lead to suboptimal transmission networks. This study investigates the potential benefits of cooperative transmission expansion planning (i.e., both intraregional and interregional planning that optimizes transmission investments across the entire physical system). Using sectoral and economic optimization, and machine learning models, it analyzes the impact of different levels of cooperation among transmission planning regions within U.S. Western Interconnection in 2019 and 2059 via an iterative investment process. Furthermore, it examines the effects of future climate change on transmission cooperation by simulating historical heat waves from 2019 under conditions of 2059. The results indicate that cooperative transmission planning leads to lower wholesale electricity prices, decreased energy outages, and reduced greenhouse gas emissions. However, the advantages of collaboration diminish during widespread heat waves, despite remaining beneficial especially for regions like California Independent System Operator with substantial solar installations. The study underscores the importance of transmission cooperation in reducing costs and enhancing reliability, emphasizing the need for strategic investments in storage to address challenges posed by future extreme weather events with varying spatial scales.

3 - Grid reliability and competing claims for scarce natural gas during winter storms

Chiara Lo Prete, Pennsylvania State University, University Park, PA, United States, Ashish Radhakrishnan

The increasing dependence of the electric power sector on natural gas poses vulnerabilities caused by competition for scarce pipeline capacity that are especially pronounced when both heating and electricity needs spike. During extreme cold weather events, pipeline deliveries to distribution companies serving heating customers receive the highest priority of service, while natural gas-fired generators may have their service curtailed regardless of contractual priority. This creates grid reliability challenges and adverse consequences that tend to fall on historically overburdened and underserved communities. In this paper, we quantify the benefits of reallocating gas consumption from industrial customers that hold firm contracts but do not serve human needs to power plants. We adopt a gas-electric optimization framework that models the fuel supply delivery risk at gas-fired power plants in the Northeastern U.S., and evaluate impacts on costs, emissions and unserved energy under alternate gas allocations during the 2014 Polar Vortex.

4 - Joint Expansion of Power and Water Distribution Networks

Sai Krishna Kanth Hari, Los Alamos National Laboratory, Los Alamos, NM, United States

Power and water distribution networks are tightly coupled in their operations. However, their expansion is planned separately. A joint expansion planning can offer benefits to both networks. To understand these benefits, we formulate the joint planning problem as a Mixed Integer Nonlinear Program. Using this formulation, we will discuss the potential benefits of joint planning and the challenges to overcome to implement it at a large scale.

WC57

Summit - Terrace Suite 1

Optimizing Decision Making in Healthcare Delivery

Invited Session

Health Applications Society

Chair: Feifan Wang, Tsinghua University, Beijing, N/A

Co-Chair: Feng Ju, Arizona State University, Tempe, AZ, United States

1 - Predicting the Number of Open Operating Rooms to Enhance Resource Allocation and Patient Quality of Care

Narges Shahraki, Mayo Clinic, Rochester, MN, United States, Nageswar Madde, Ryan Schoer, Ellory Steinbauer, Aimee Tillman, Matthew Vogt, Daryl Kor

The dynamic nature of surgical demand poses challenges in predicting the number of open Operating Rooms (ORs), which may lead to inefficient resource allocation and potential negative impacts on patient quality of care. This study introduces a novel deep learning framework with a wide and deep architecture to predict the daily maximum number of open ORs. The daily maximum number of open ORs is a crucial metric for optimal recourse allocation of personnel. The predictive framework can improve resource allocation strategies, thereby enhancing patient quality of care by minimizing delays in providing care and reducing staff burnout.

2 - Patient Level Risk Prediction and Management in ICU

Hangtian Li, Tsinghua University, Beijing, China, People's Republic of, Xiaolei Xie, Feng Ju

Intensive Care Units (ICU) are associated with substantial costs, while they generate extensive patient-level physiological monitoring data. With the development of electronic medical records and data mining technology, we can use these data to predict the development of patients' conditions and then optimize the allocation of medical resources to improve patient care quality, ICU operational efficiency, and resource utilization. Based on public critical medical databases, we propose a multimodal machine learning model to model and predict patients' disease progression and risk. The proposed predictive model not only

integrates multimodal patient data for prediction but also adeptly captures correlations between each modality. Specifically, it can effectively identify the relationship between patient physiological monitoring time series data and their diagnosis/treatment procedures, thus enhancing prediction accuracy and reliability. The predictive model is based on a graph structure and can make recommendations for subsequent clinical examinations, which increases the model's interpretability and clinical utility. To validate our model, we trained the model on a large dataset to predict changes in multiple physiological indicators at the patient level. Furthermore, through the integration of patient-level modeling and predictive analytics with operations research algorithms, we optimize ICU medical resource scheduling, ultimately enhancing ICU operation quality and efficiency.

3 - Leveraging Wearable Technology to Improve Fall Prevention: Early Lessons Learned

Emma Fortune, Mayo Clinic, Rochester, MN, United States, Terri Menser, Che Ngufor, Louis Faust, Philip Tipton

Falls account for 87% of fractures in the elderly, incurring annual costs of nearly $50 billion in the U.S. Clinical assessments of fall risk are reported to have low sensitivities, provide an incomplete picture of a patient's functional level, and introduce access barriers. Remote monitoring technologies including smart phones and watches have shown promise with real-world gait and stability measurement but are not widely adopted in healthcare settings; secure and reliable data transmission remains a main challenge.

Our goal is to develop a fall prediction tool by incorporating available objective daily physical activity (PA), stability, and/or cardiopulmonary health data from patients' existing smart phones and watches with typical clinical risk factors from the electronic health record (EHR), which can provide standalone generalizability for widescale deployment. We will recruit 200 Mayo Clinic patients aged >64 years and collect pilot data for initial model and implementation strategy development. Eligibility criteria include owning an iPhone, and able to walk >1 block unassisted.

Participants will be asked to share their daily PA, stability, and/or cardiopulmonary data from their Apple device(s) for 6-months. Falls will be tracked during those 6-months using EHR-reported falls data, weekly fall survey, and the Apple Fall Detection event log. An AI predictive model will be developed using the gait and clinical data as inputs and validated using the combined fall tracking data as ground truth. The sensitivities and specificities of our model and the STEADI survey will be compared. Our implementation strategy will be refined by early lessons learned.

4 - Optimizing Technician Staffing in Medical Services Using Two-Stage Stochastic Integer Programming

Mirui Zhang, Tsinghua University, Beijing, China, People's Republic of, Feifan Wang

The technician in medical services is an important role in supporting medical procedures running smoothly. Due to the uncertainty in technician workload, technicians may often have to work overtime. It is widely seen in procedure rooms and operating rooms and may compromise patient care quality and technician job satisfaction. In this study, we focus on technician staffing, in combination with scheduling. We develop a two-stage stochastic integer programming model to improve workload coverage. Multiple scheduling scenarios are considered to determine the optimal configuration of a technician team. The proposed method is evaluated through numerical experiments. The results show that the method can significantly reduce overtime and labor cost.

WC58

Summit - Terrace Suite 2

Advances in Healthcare Analytics

Invited Session

Health Applications Society

Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

Co-Chair: Vasileios Digalakis, HEC Paris, Jouy-en-Josas, France

1 - Analyzing Game Play Patterns and Their Association with Health Knowledge

Rema Padman, Carnegie Mellon University, Pittsburgh, PA, United States, Rahul Ladhania, Jeong Hin Chin, Abhishek Gupta

With the widespread mobile device penetration in society and easy access, gamification can be an engaging way for children to learn about healthy lifestyle related habits early in life and improve lifelong health risks. This highlights the promise of a 'digital vaccine' framework, a subcategory of digital therapeutics, which leverage AI and mobile technology-enabled solutions to reduce disease risk and improve health and well-being. In this study, we use data from a novel, 11-week, cluster randomized controlled trial we conducted in India to evaluate the impact of regular exposure in the school setting to an artificial intelligence(AI)-enabled, educational mobile health game – a low risk, non-invasive, digital vaccine candidate with neurocognitive training and implicit learning components – on students' health related knowledge. We present results examining gameplay patterns of students who were exposed to the mHealth game, and to quantify association with their health related knowledge, comprising nutrition, physical activity, and health hygiene habits, captured from student surveys over the course of the intervention. Early results from supervised and unsupervised learning of children's play patterns and their relationship with their "health knowledge" suggests the role of gameplay features such as performance on the game (win rate) and repeated exposure (number of games played) in predicting changes in participants' learning and knowledge. Examining the patterns over time also show significant variations among participants, with two types of gameplay behavior. Ongoing analyses will provide new insights into the complex interactions between gameplay and health behavior changes for designing more impactful interventions.

2 - A Game-Theoretic Framework of Fairness Dynamics: Equilibrium, Control and Beyond

Chuwen Zhang, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Pengyi Shi, Amy Ward

Recidivism control is a crucial aspect of modern criminology. Recently, there is a growing interest in the use of algorithmic sentencing, where the jury uses a risk estimate to make incarceration decisions. However, this approach often incorporates attributes beyond criminal recosrds, such as race, education, and unemployment, leading to significant fairness issues. We introduce a queuing-type system to model the crime justice system that delineates the dynamics of population and evolution of offending probability, in which the recidivists are modeled as a party of antagonistic players against the system, while the latter reacts by placing control policies that both affect the incentive and deterrence for recidivists. Thereby, we introduce a utility function that mimics the tradeoff of redivists. We capitalize on the theory of the generalized Nash game to facilitate a discussion on the existence of the equilibria and convergence therein. We start by the basic existence properties under fixed control policies. The uniqueness of equilibria can be captured by a curvature condition, which indicates the sufficiency of deterrence. Based on the basic stationary controls, we discuss further investigations in algorithmic sentencing settings and extensions with multiple demographic groups, in which we discuss the issues of fairness and disparity. These results provide insights on when and nor fairness constraints may be effective.

3 - Exact Sensitivity Analysis of Markov Reward Processes via Algebraic Geometry

Muhammad Maaz, University of Toronto, Toronto, ON, Canada, Timothy Chan

Markov chains are commonly used for economic modelling, especially in health economics. The outcomes of health economic analyses are functions of the primitives of the Markov chain. As these primitives are often uncertain, it is standard to do a sensitivity analysis, which is typically performed as a grid search in the parameter space. In this work, we introduce a rigorous method for sensitivity analysis of Markov models. Our key insight is that all typical health economic analyses can be reformulated as systems of polynomial inequalities. Solving this system gives an exact characterization of the space of parameters that yield the modeller's desired outcome. The algorithm of \textit{cylindrical algebraic decomposition} (CAD) is the usual method to solve polynomial systems, but it takes doubly exponential time, which is intractable for more than a few variables. We introduce a special class of polynomial systems consisting of a single defining inequality and a series of simplex constraints. Using tools from algebraic geometry, we introduce a novel class of polynomials which we call \textit{simplex-extensible}, and show that if the defining inequality consists of such a polynomial, then we can run CAD in singly exponential time. Importantly, the polynomial systems induced by Markov modelling are instances of this special system and have the simplex-extensible property. Using our theoretical results, we implement a software package which can symbolically manipulate Markov chains and construct CADs of inequalities arising from them.

4 - Learning Adaptive Behavioral Interventions to Reduce Ascvd Risk

John Silberholz, University of Michigan Ross School of Business, Ann Arbor, MI, United States, Hamsa Bastani

Increasing patient engagement in chronic disease prevention and management behaviors is crucial to the success of population health and value-based care programs, and many behavioral interventions have been studied for this purpose. A common challenge with these interventions is that they are often static -- they do not evolve based on user engagement, success, or failure. As a result, they may allocate resources in inefficient ways, such as using significant time of personnel to contact patients who are low risk or who are unengaged and unlikely to benefit from further engagement. This can be problematic, as behavioral interventions can be associated with significant costs. In this work, we develop a dynamically adapted intervention to engage patients in preventive health behaviors for Atherosclerotic Cardiovascular Disease (ASCVD), the leading cause of mortality worldwide. A key requirement for a dynamically adaptive algorithm is that it quickly identifies an effective approach to targeting patients with an intervention. To achieve this goal, we use data from five past ASCVD behavioral interventions to “warm-start” a proposed dynamically adapted intervention. We discuss ongoing work to pilot the proposed intervention.

WC61

Summit - Ballroom 3

Reinforcement Learning

Invited Session

Applied Probability Society

Chair: Qiaomin Xie, University of Wisconsin-Madison, Madison, United States

1 - Approximate Global Convergence of Independent Learning in Multi-Agent Systems

Zaiwei Chen, Purdue University, West Lafayette, IN, United States, Ruiyang Jin, Yiheng Lin, Jie Song, Adam Wierman

Independent learning (IL), despite being a popular approach in practice to achieve scalability in large-scale multi-agent systems, usually lacks global convergence guarantees. In this paper, we study two representative algorithms, independent Q-learning and independent natural actor-critic, within value-based and policy-based frameworks, and provide the first finite-sample analysis for approximate global convergence. The results imply a sample complexity of $\tilde{\mathcal{O}}(\epsilon^{-2})$ up to an error term that captures the dependence among agents and characterizes the fundamental limit of IL in achieving global convergence. To establish the result, we develop a novel approach for analyzing IL by constructing a separable Markov decision process (MDP) for convergence analysis and then bounding the gap due to model difference between the separable MDP and the original one. Moreover, we conduct numerical experiments using a synthetic MDP and an electric vehicle charging example to demonstrate our results and the practical applicability of IL.

2 - Rectangularity and Duality of Distributionally Robust Markov Decision Processes

Yan Li, ISyE Georgia Tech, Atlanta, GA, United States, Alexander Shapiro

The main goal of this paper is to discuss several approaches to the formulation of distributionally robust counterparts of Markov Decision Processes, where the transition kernels are not specified exactly but rather are assumed to be elements of the corresponding ambiguity sets. The intent is to clarify some connections between the game and static formulations of distributionally robust MDPs, and delineate the role of rectangularity associated with ambiguity sets in determining these connections.

3 - Distributionally Robust Reinforcement Learning with Interactive Data Collection: Fundamental Hardness and Near-Optimal Algorithm

Miao Lu, Stanford University, Stanford, CA, United States, Han Zhong, Tong Zhang, Jose Blanchet

The sim-to-real gap, which represents the disparity between training and testing environments, poses a significant challenge in reinforcement learning (RL). A promising approach to addressing this challenge is distributionally robust RL, often framed as a robust Markov decision process (RMDP). In this framework, the objective is to find a robust policy that achieves good performance under the worst-case scenario among all environments within a pre-specified uncertainty set centered around the training environment. Unlike previous work, which relies on a generative model or a pre-collected offline dataset enjoying good coverage of the deployment environment, we tackle robust RL via interactive data collection, where the learner interacts with the training environment only and refines the policy through trial and error. In this robust RL paradigm, two main challenges emerge: managing distributional robustness while striking a balance between exploration and exploitation during data collection. Initially, we establish that sample-efficient learning without additional assumptions is unattainable owing to the curse of support shift. To circumvent such a hardness result, we introduce the vanishing minimal value assumption to RMDPs with a total-variation (TV) distance robust set, postulating that the minimal value of the optimal robust value function is zero. We prove that such an assumption effectively eliminates the support shift issue for RMDPs with a TV distance robust set, and present an algorithm with a provable sample complexity guarantee. Our work makes the initial step to uncovering the inherent difficulty of robust RL via interactive data collection and sufficient conditions for designing a sample-efficient algorithm.

4 - Value-Incentivized Preference Optimization: A Unified Approach to Online and Offline RLHF

Shicong Cen, Carnegie Mellon University, Pittsburgh, PA, United States, Jincheng Mei, Katayoon Goshvadi, Hanjun Dai, Tong Yang, Sherry Yang, Dale Schuurmans, Yuejie Chi, Bo Dai

We demonstrate that principled exploration for reinforcement learning (RL) can be achieved for general policy parameterizations, including large language models (LLMs), without requiring explicit uncertainty modeling. Avoiding explicit uncertainty modeling is critical in real world RL applications, such as preference alignment of LLMs using reinforcement learning from human feedback (RLHF) or direct preference optimization (DPO), since conducting uncertainty calculations over arbitrary nonlinear models is intractable. Our proposed solution, Value-Incentivized Preference Optimization (VIPO), utilizes an easy to calculate objective that admits theoretical guarantees. Experiments on text summarization and dialog verify the practicality and effectiveness of the proposed approach.

WC62

Summit - Signature Room

Recent Development in Revenue Management and Supply Chain Management

Invited Session

Revenue Management and Pricing

Chair: Anyan Qi, The University of Texas at Dallas, Richardson, TX, United States

1 - Component Pricing with a Bundle Size Discount

Zechao Li, Tsinghua University, Beijing, China, People's Republic of, Ningyuan Chen, Xiaobo Li, Chun Wang

Firms selling multiple products usually adopt bundle pricing in their marketing strategy for the purpose of extracting large consumer surplus. In this paper, we propose and analyze a bundling mechanism, referred to as component pricing with a bundle size discount (CPBSD), which sells bundles for the summed prices of the included products (component pricing) minus a discount based on the number of products purchased (bundle size discount). CPBSD is conceptually simple and has been widely used in real-world business. Theoretically, we show that CPBSD subsumes several well-studied bundling mechanisms as special cases. Moreover, we find instances in which other common mechanisms only generate a fraction of the optimal profit of CPBSD. We further prove that, under a general condition, CPBSD attains the optimal profit asymptotically for a large number of products among all bundle pricing mechanisms. Practically, we formulate a mixed-integer linear program for the optimal pricing scheme of CPBSD, and we also develop an approximation algorithm for efficiently solving CPBSD in large-scale problems. Through comprehensive numerical experiments, we show that CPBSD demonstrates superior performance in contrast to other bundling mechanisms. In particular, compared to bundle size pricing whose outstanding empirical performance has been extensively tested in the literature, CPBSD performs especially well when products are heterogeneous and the production costs are high. Furthermore, the outstanding performance of CPBSD is most pronounced when the potential surplus provided by products are negatively correlated with the product valuations.

2 - A Prior-Free, Asymptotically Efficient Mechanism for Two-Sided Platforms with Dynamic Arrivals

Arun Kumar Rout, The University of Texas at Dallas, Richardson, TX, United States, Milind Dawande, Ganesh Janakiraman

We consider a platform that operates a two-sided market in which unit-demand buyers with private valuations and unit-supply sellers with private costs arrive dynamically over time -- the arrival processes of the agents (buyers and sellers) and their value/cost distributions can be quite general in that these inputs are allowed to be Markov-modulated. The platform incurs a waiting cost per agent per unit time for unmatched agents waiting in the market. The platform wishes to design a mechanism that specifies how and when agents are matched or dismissed, subject to a budget-balance constraint. The platform's objective is to maximize the efficiency of the mechanism, defined as the difference of the aggregate valuation of the matched buyers and the aggregate cost of the matched sellers, less the waiting cost. For this problem, we develop a prior-free and dominant-strategy incentive compatible mechanism. For any choice of problem parameters, we derive a closed-form upper bound on the absolute efficiency gap of this mechanism. In addition, we show that this mechanism is asymptotically efficient (i.e., the relative efficiency gap of our mechanism approaches zero as the market size increases). More generally, our analysis offers a sophisticated yet intuitive framework that accepts as input a static mechanism (where all information is static) with some desirable properties and outputs a dynamic mechanism that retains these properties and is also prior-free and asymptotically efficient.

3 - Beyond One-Size-Fits-All: Personalized Delivery and Fulfillment Optimization

Quan Zhou, McGill University, Montreal, QC, Canada, Mehmet Gumus, Sentao Miao

Problem definition: Motivated by our collaboration with an online platform operating in North America, we explore the joint optimization of the order fulfillment process with personalized delivery options in the context of ecommerce. Each customer can choose from a personalized set of fulfillment options to proceed with the purchase or leave with no purchase. Fulfillment assignments of purchased orders are determined periodically. **Methodology**: We model customer behavior with a general discrete choice model and formulate the joint optimization as a stochastic dynamic program. We propose a tractable deterministic approximation and develop a computationally-efficient heuristic with a provable performance guarantee. We also extend the proposed our heuristic to the price-sensitive customer case. **Managerial implications**: By using both synthetic and real datasets collected from our industrial partner, we demonstrate the value of personalizing fulfillment options for the customers and jointly optimizing the options with fulfillment assignments. Our results show that demand management via personalized fulfillment options is prominent when customers favor quicker fulfillment and/or the fulfillment capacity is limited. However, an optimized fulfillment operation is more pivotal when customers are more willing to wait.

4 - Why Perfect Machine May Not be Perfect: Incentivizing Human-Machine Collaboration

Xiaotong Guan, Fudan University, Shanghai, China, People's Republic of, Anyan Qi, Shouqiang Wang

Humans and machines need to work collaboratively to make operational decisions such as demand forecasts. In this paper, we study how the machine should be designed to incentivize human-machine collaboration and obtain a rich set of results.

5 - Inventory Allocation Under the Greedy Fulfillment Policy: the (Potential) Perils of the Hindsight Approach

Sheng Liu, University of Toronto, Toronto, ON, Canada, Stefanus Jasin, Jinglong Zhao

We consider an online retailer with multiple warehouses and multiple demand locations. We assume that the retailer uses the greedy fulfillment policy (i.e., ship from the closest and/or cheapest warehouse) and consider the performance of the initial inventory allocation given by the hindsight or the stochastic programming approach, under which the retailer can first observe all the realized demands before simultaneously making the fulfillment decisions to minimize the total cost. The hindsight approach has been proposed by recent works in the literature and is practically appealing. While this approach has its merits, its limitations are not well understood. In this work, we first show that the hindsight solution is equivalent to the optimal inventory allocation under the greedy fulfillment policy but on a misspecified (more optimistic) demand arrival distribution. This optimism is sometimes benign but can also backfire. We provide necessary and sufficient conditions for the hindsight solution to be asymptotically optimal (as the unit lost sale cost grows large) under the greedy fulfillment policy. If any of these conditions are violated, we show that the expected total cost under the hindsight solution and the greedy fulfillment policy can be arbitrarily bad compared to the optimal expected total cost under the greedy fulfillment policy. Finally, we show that even if we allow the retailer to use the best fulfillment policy once the hindsight solution has been computed, in general, its expected total cost can still be arbitrarily bad compared to the optimal expected total cost under the greedy fulfillment policy.